LWG
LOWER WILLAMETTE GROUP

Chairperson: Bob Wyatt, NW Natural
Treasurer: Frederick Wolf, DBA, Legacy Site Services for Arkema

Via E-mail & Hand Delivery

August 26, 2014

Richard Albright, ECL Director
United States Environmental Protection Agency, Region 10
Office of Environmental Cleanup, Mail Code ECL-117
1200 Sixth Avenue, Suite 900
Seattle, Washington 98101-3140

> **Re: Request for Dispute Resolution of EPA's Notice of Decisions on Background Regarding Section 7 of the Remedial Investigation; Lower Willamette River, Portland Harbor Superfund Site, USEPA Docket No: CERCLA-10-2001-0240**

Dear Mr. Albright:

The Lower Willamette Group ("LWG") objects to the U.S. Environmental Protection Agency's ("EPA") decisions on background in Section 7 of the Remedial Investigation ("RI") for the Portland Harbor Superfund Site ("Site"). These decisions are stated in the August 12, 2014 e-mail from Deb Yamamoto (copy enclosed). The LWG requests resolution of this dispute in accordance with Section XVIII of the Administrative Settlement and Order on Consent for the RI/FS ("AOC") for the Site. This dispute resolution request was timely filed.

Summary of Dispute and Requested Relief

The LWG objects to EPA's approach to defining background data for upriver bedded sediment. EPA's approach errs in two fundamental ways. First, contrary to EPA guidance, EPA discards analytically valid environmental data. The guidance (USEPA 2009b, p.6-36) states: "If no error in the value can be documented, it should be assumed that the observation is a true but extreme value. In this case, it should not be altered or removed." Second, in choosing which data to declare as outliers, EPA makes numerous statistical errors, which render its decision to exclude the data from its background calculations erroneous and an abuse of discretion. However, even if these statistical errors were ignored, it is a fact that upstream bedded sediments with elevated concentrations may be transported downstream to the Site. A statistical categorization of the data, right or wrong, does not affect the possibility or likelihood of this transport.

The calculation of background bedded sediment concentrations is not an academic or statistical exercise, but rather a very real, tangible estimate of mobile upstream contaminant mass. A gross underestimation of the potentially mobile upstream bedded sediment contaminant

mass is likely to result in a Site remedy with unachievable remedial action goals, which runs contrary to EPA guidance.

The following example illustrates the purpose of calculating background bedded sediment concentrations. Imagine a flood that scours one foot of sediment from the entire upstream reach of the Willamette River that was used to calculate background concentrations. That mobile sediment is then homogenized in the water column as it is carried downstream. Some of that homogenized sediment is deposited in certain areas within the Superfund Site. The question that the background sediment calculation aims to answer is this: What is the best estimate for the concentration of contaminants in that mobile, homogenized sediment?

The only way to accurately estimate the mobile, homogenized sediment concentration is to accurately estimate the mass of contaminants in the upstream bedded sediment, regardless of its location or concentration. All valid data must be used to make that estimate accurately, because all bedded upstream sediment could potentially be mobilized and carried downstream. The statistical distribution of the data for upstream bedded sediment concentrations, or the presence of statistical "outliers" in the data set, are irrelevant to the proper estimate of potentially mobile contaminant mass in the upstream bedded sediment.

The LWG objects to EPA's approach to defining background data for upriver bedded sediment for the following specific reasons:

1. EPA abused its discretion by excluding outliers from the reference area data set. EPA guidance states: "One should never discard an outlier based solely on a statistical test. Instead, the decision to discard an outlier should be based on some scientific or quality assurance basis." (EPA 2000a). EPA did not scientifically assess the outliers (e.g., weigh all available lines of evidence) to determine the reason for the elevated values, which is essential to determine whether they should be retained or removed.

2. EPA abused its discretion (and erred scientifically) when it calculated upriver sediment concentrations using tests to identify outliers that explicitly assume a normal distribution for all populations from which the sample data were obtained.

3. EPA abused its discretion by arbitrarily setting the number of suspected outliers to 10 for all outlier tests it performed, contrary to the advice of EPA guidance documents, which recommend using graphical techniques to determine the number of potential outliers for testing.

4. Contrary to EPA guidance documents, EPA discarded observations statistically identified as outliers based on an improperly applied statistical test without investigating whether any evidence justified discarding those observations, such as analytical quality issues or site-specific environmental conditions.

5. EPA failed to use correct statistical methods to evaluate Q-Q plots or to otherwise formally test for outliers in datasets that contain nondetect values ("NDs"), such as the use of Tobit regression.

6. EPA's justification for removing "outliers" is based on its concept that "reference area data may also contain high-biasing outliers that are either not representative of the dominant background population or are representative of specific contaminant sources."

2

RI, Section 7.3. However, because upstream bedded sediments with elevated concentrations may be transported downstream to the Site, it is important for the reference area data to represent the **total** reference area population, not a post-hoc background population constructed by the removal of valid data.

These errors are based on either a failure to follow EPA guidance or the application of that guidance in a way that is inconsistent with standard statistical approaches to analyzing environmental data. Consequently, EPA exceeded its discretion. These errors will have a significant effect on the understanding of the Site, as well as on the assessment of remedial technologies and remedial decision making. Upstream data values are used in developing surface weighted average concentrations ("SWAC") for the background area. Because the area upstream of the Site is a source of suspended sediment that will be transported to and deposited within the Site, knowledge of upstream bedded sediment concentrations is an important prerequisite value in the calculation of Site equilibrium concentrations. Site equilibrium values are limiting factors in the assessment of the effectiveness of remedial technologies, in determining residual risk levels, and in evaluating Site recovery.

According to EPA guidance, when sediment concentration values that have been determined to be analytically valid and not otherwise in error (e.g., due to an analytical data quality issue or a transcription error) are improperly deleted from a data set, important site-specific information is lost (USEPA 2006a). For example, when the highest values are excluded from the data set, the resulting estimated SWAC value is erroneously low and not representative of actual site conditions. A critical error is then created when the background data set is used in the RI to evaluate whether site-specific releases in the study area have resulted in elevated chemical concentrations, and in the FS when the data are used to estimate the sediment concentrations that are achievable for the remediation alternatives. This error then propagates throughout the FS, resulting in a cascade of errors and inaccuracies related to:

- The evaluation of effectiveness of remedial technologies;
- The determination of residual risk levels; and
- The evaluation of natural recovery.

EPA guidance acknowledges the importance of assessing anthropogenic background, including whether releases (historical or current) in the upstream reference area that are unrelated to the CERCLA site justify retention of elevated background concentrations in the calculation of background threshold values ("BTVs"):

"Where background concentrations are high relative to the concentrations of released hazardous substances, pollutants, and contaminants, a comparison of site and background concentrations may help risk managers make decisions concerning appropriate remedial actions. The contribution of background concentrations to risks associated with CERCLA releases may be important for refining specific cleanup levels for COCs that warrant remedial action." (USEPA 2002a, p. B-6).

"It is especially important to consider both background levels of contamination and what has been achieved at similar sites elsewhere, so that achievable cleanup

levels are developed. All of these factors should be considered when establishing final cleanup levels that are within the risk range." (USEPA 2005, p. 2-17).

"The project team and site experts should decide what represents a site population and what represents a background population. The project team should determine the population area and boundaries based upon all current and future uses, and the objectives of data collection." (USEPA 2013, p.23).

Additionally, when assessing anthropogenic background, locations and characteristics of the sampling stations used in the background area should reflect the physical, chemical, geological, and biological characteristics of the site itself. Here, the reference area samples exhibiting the highest PCB concentrations were in areas, similar to the Site, whose physical and hydrological characteristics cause them to be depositional and where the sediments that have been deposited exhibit fine grain sizes and organic carbon content that best reflect the sediments encountered in the majority of the Study Area. The reference area samples with lower concentrations were obtained in areas of known scour and with much lower percent fines and organic material, and therefore do not as closely resemble the type of upstream sediments that are likely to be transported and deposited in the Site.

Thus, by excluding samples with the highest concentrations of Total PCBs, EPA removed data that best represent the background sediment contamination concentrations upstream of the Site that are most likely to be transported downstream and deposited into the Site.

Retaining the appropriate samples for the calculation of BTVs is also important to provide EPA risk managers with key information to help develop and select remediation goals, and for risk communicators to convey information to the public. In particular, EPA should not assume elevated background concentrations will be remediated, and therefore dropped from the calculation of BTVs:

"The presence of high background concentrations of COPCs may pose challenges for risk communication. For example, *the discussion of background may raise the expectation that EPA will address those risks under CERCLA.* The knowledge that background substances may pose health or environmental risks could compound public concerns in some situations.

On the other hand, knowledge of background risks could help some community members place CERCLA risks in perspective. Also, *the information about site and background risks can be helpful for both risk managers who make an appropriate CERCLA decision, and for members of the public who should know about environmental risk factors that come to light during the remedial investigation process*." (USEPA 2002b, p. 8) (emphasis added).

" . . . project managers should understand the role of the contaminated water body in the watershed, including the habitat or flood control functions it may serve, *the presence of non-site-related contaminant sources in the watershed* . . . . In these areas, it can be especially important to consider background concentrations when

developing remedial objectives and to evaluate the incremental improvement to the environment if an action is taken at a specific site in the watershed." (USEPA 2005, p. 2-17) (emphasis added).

EPA guidance also cautions against discarding background samples that exhibit high concentrations (e.g., concentrations that may statistically *appear* as outliers):

"One should never discard an outlier based solely on a statistical test. Instead, the decision to discard an outlier should be based on some scientific or quality assurance basis. Discarding an outlier from a data set should be done with extreme caution, particularly for environmental data sets, which often contain legitimate extreme values. If an outlier is discarded from the data set, all statistical analysis of the data should be applied to both the full and truncated data set so that the effect of discarding observations may be assessed. If scientific reasoning does not explain the outlier, it should not be discarded from the data set." (USEPA 2006a, p. 51; USEPA 2006b, p.116).

EPA included the following statement in its revisions to Section 7 to justify discounting outliers:

"Although it is not necessary for the data to be normally distributed to apply either Dixon's or Rosner's test, the resulting data after the potential outliers are removed should follow a normal distribution. However, this condition was not met in all instances, and thus ***greater emphasis was given to the visual examination of the data to supplant the results of the statistical tests alone***."

RI, Section 7.3 (emphasis added).

By not requiring consideration of site-specific scientific information in order to determine whether an outlier should be discarded, EPA failed to follow its own guidance. EPA should have considered other key lines of site-specific scientific evidence (e.g., hydrodynamics, watershed and atmospheric sources, grain size and organic carbon content of the sampled sediment), instead of relying on "visual examination of the data."

EPA also erred in discarding the outliers because these samples best represent the physical conditions of the majority of the Study Area (i.e., in terms of percent fines and percent organic carbon). EPA guidance states:

"Generally, the type of background substance (natural or anthropogenic) does not influence the statistical or technical method used to characterize background concentrations. For comparison purposes soil samples should have the same basic characteristics as the site sample (i.e., similar soil depths and soil types)." (USEPA 2002b).

The samples from the upstream reference area that EPA retained in calculating background values exhibited lower concentrations of contaminants and were obtained from areas of known scour (i.e., with much lower percentages of fines and organic carbon).

The LWG therefore requests that the full data set with consideration of organic carbon correction be retained as the selected set of background values and applied in the FS. This data set is the most appropriate one for determining BTVs. These values are shown in the "all data" columns of Table 7.3-1b (and the related Appendix H Table H-2b) of the RI Section 7 revision agreed to by EPA and the LWG on December 12, 2013. This approach is consistent with EPA guidance and the NCP, it is based on sound science, and it is supported by the other strong lines of evidence presented in the RI.

Dr. Steve Millard, an independent statistical consultant, has assisted the LWG in analyzing EPA's approach, and he is available to meet with you to answer any questions. Dr. Millard is the principal at Probability, Statistics & Information ("PSI"), as well as a biostatistician at the VA Puget Sound Health Care System in Seattle, Washington, and has worked in the field of environmental and health care statistics for over 30 years. He is an author and co-author of textbooks on environmental statistics and statistics for drug development, and the creator of the R package *EnvStats*. Dr. Millard holds a B.A. in Mathematics from Pomona College, and an M.S. and Ph.D. in Biostatistics from the University of Washington.

Discussion[1]

Portland Harbor is affected by sources of ***upstream*** contamination that are unrelated to CERCLA releases. The RI addresses this reality by comparing chemical concentrations in the "remediated" site with concentrations in the "background" or "reference" area. If other sources of contamination will affect the remediated site even after cleanup, then the reference area needs to reflect these additional sources. EPA guidance states "[a] background reference area should have the same physical, chemical, geological, and biological characteristics as the site being investigated, but has not been affected by activities on the site." USEPA (2002a, p. 2-2). EPA guidance further states that soil samples in the reference area "should have the same basic characteristics as the site sample [*sic*] (i.e., similar soil depths and soil types)." USEPA (2002a, p. 1-2).

When computing BTVs based on data from the reference area, it is contrary to standard practice and sound science to exclude high chemical concentrations based upon a suspicion of an unknown or unsubstantiated source, or even a source that has been identified but is not likely to be remediated. Instead, EPA guidance states that "[i]f scientific reasoning does not explain the

---

[1] The following three EPA guidance documents focus on statistical models for environmental data and are referenced throughout this letter and will sometimes be referred to as "the three EPA guidance documents":

- *ProUCL 5.0.00 Technical Guide* (USEPA, 2013);
- *Scout 2008 Version 1.0 User Guide* (USEPA, 2009a); and
- *Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities: Unified Guidance* (USEPA, 2009b).

outlier, it should not be discarded from the data set," (USEPA 2006a, p. 51; USEPA 2006b, p.116), and also provides that the following procedure should be used:

> "If an error in transcription, dilution, analytical procedure, etc. can be identified and the correct value recovered, the observation should be replaced by its corrected value and further statistical analysis done with the corrected value.
>
> If it can shown [*sic*] that the observation is in error but the correct value cannot be determined, the observation should be removed from the data set and further statistical analysis performed on the reduced data set. The fact that the observation was removed and the reason for its removal should be documented when reporting results of the analysis.
>
> ***If no error in the value can be documented, it should be assumed that the observation is a true but extreme value. In this case, it should not be altered or removed. However, if feasible, it may helpful [*sic*] to obtain another observation in order to verify or confirm the initial measurement***."

USEPA (2009b, pp. 6-35, 6-36) (emphasis added). EPA contends that its handling of the values is consistent with guidance, but the record contains no evidence that it followed this procedure.

Critically, whether or not an observation appears to be a "large" value compared to the rest of the observations usually depends on the statistical model chosen. All three EPA guidance documents that discuss statistical models (USEPA 2009a,b; 2013) state that environmental data may be modeled by the normal distribution, lognormal distribution, gamma distribution, or some other distribution, and they suggest using graphical techniques such as Q-Q plots and goodness-of-fit ("GOF") tests to determine the appropriate distribution to use. Instead, EPA used Rosner's test (which assumes the underlying data are normally distributed, except perhaps a pre-specified number of outliers) with the pre-specified number of outliers set to 10 to test for outliers, and also looked at normal (Gaussian) Q-Q plots. In essence, EPA assumed that all underlying COC distributions are or should be normal (Gaussian).

EPA's decision to exclude outliers is contrary to EPA guidance and standard approaches to analyzing environmental data in the following ways:

Standard Approach	EPA Approach
Decide what distribution applies to the data.	Implicitly assumed the background data, except for possible outliers, was normally distributed.
EPA guidance recommends using graphical techniques to determine the number of outliers to test for.	EPA arbitrarily set the number of suspected outliers to 10 for all outlier tests and then removed all statistical outliers identified.
Do not discard data based solely on a test for outliers.	Discarded data observations determined to be outliers without investigation whether or not any evidence justifies discarding the observations.

Use the correct procedures when the data contain ND values.	EPA did not use correct statistical methods to properly use Q-Q plots or to formally test for outliers for datasets that contain NDs.

These errors result in a misunderstanding of Site characteristics that EPA has already decided to carry forward into the FS, resulting in an adverse impact on the evaluation of the effectiveness of remedial technologies, the determination of residual risk levels, and the evaluation of natural recovery.

1.   EPA Abused Its Discretion by Assuming Normal Distribution and Not Properly Deciding What Distribution to Use

All three EPA guidance documents (USEPA 2009a,b; 2013) discuss the importance of identifying whether environmental data sets are properly modeled by normal distribution, lognormal distribution, gamma distribution, or some other distribution:

> "Many environmental data sets can be modeled by a gamma as well as a lognormal distribution." (USEPA 2013, p.1).

> "In practice, many skewed data sets can be modeled both by a lognormal distribution and a gamma distribution. . . . It is suggested that all skewed data sets be first tested for a gamma distribution." (USEPA 2013, pp. 77, 79).

> "Consequently, an important facet of choosing among appropriate test methods is determining whether a commonly-used statistical distribution such as the normal, adequately models the observed sample data. A large variety of possible distributional models exist in the statistical literature; most are not typically applied to groundwater measurements and often introduce additional statistical or mathematical complexity in working with them. So groundwater statistical models are usually confined to the gamma distribution, the Weibull distribution, or distributions that are normal or can be normalized via a transformation (e.g., the logarithmic or square root). . . .

> "Assumptions of normality are most easily made with regard to naturally-occurring and measurable inorganic parameters, particularly under background conditions. Many ionic and other inorganic water quality analyte measurements exhibit decent symmetry and low variability within a given well data set, making these data amenable to assumptions of normality. Less frequently detected analytes (e.g., certain colloidal trace elements) may be better fit either by a site-wide lognormal or another distribution that can be normalized, as well as evaluated with non-parametric methods." (USEPA 2009b, pp. 10-1, 10-5).

The three EPA guidance documents suggest using graphical techniques such as Q-Q plots and boxplots, as well as GOF tests and tests for outliers to determine the appropriate distribution to use, or whether to use a nonparametric method. However, EPA did not conduct any modeling

or examine the graphical techniques prepared as part of the revised RI Section 7 data products to determine the appropriate distribution. It instead assumed normal distributions.[2]

EPA may contend that this assumption is consistent with the ProUCL 5.0.00 Technical Guide (USEPA, 2013, p. 189), which also states that "[o]utlier tests should be performed on raw data, as the cleanup decision needs to be made based upon concentration values in the raw scale and not in log-scale or some other transformed scale (e.g., cube root)." Although one might infer from this statement that one should always assume the data come from a normal distribution, this statement should not be read in isolation because it is contrary to other statements in the guidance, which note that "[m]any environmental data sets can be modeled by a gamma as well as a lognormal distribution." (USEPA, 2013, p. 1). The ProUCL 5.0.00 Technical Guide contains an in-depth discussion on statistical methods for computing hypothesis tests, confidence limits, prediction limits, or tolerance limits based on assuming anything other than a normal distribution. Rather than applying the standard approach of using modeling and graphical techniques to determine distribution and applying the statistical methods appropriate for that distribution, EPA ran all tests for outliers using either Rosner's or Dixon's test on the raw data. This approach, although arguably consistent with the one above-referenced sentence in the ProUCL 5.0.00 Technical Guide (USEPA 2013), is contradictory to the guidance as a whole and to standard statistical approaches and, therefore, is an abuse of discretion. The problem that results from this approach is that the normal distribution assumption then dictates the type of statistical test models used on the data set. However, because the distribution assumption is flawed, the modeling test is applied inappropriately and improperly, resulting in deletion of sample values that are analytically valid.

2.   **EPA Abused Its Discretion by Not Using Graphical Techniques to Set the Number of Suspected Outliers for Testing**

Both USEPA (2013) and USEPA (2009a) recommend using Rosner's or Dixon's test for outliers, but because these tests assume the data without suspected outliers comes from a normal distribution, the applicability of the recommendation has to be limited to those circumstances. USEPA (2013, p. 189) states:

> "The classical outlier tests, Dixon and Rosner tests, assume that the data set without the suspected outliers follow a normal distribution; that is for both Rosner and Dixon tests, the data set representing the main body of the data obtained after removing the outliers (and not the data set with outliers) needs to follow a normal distribution."

Whether or not an observation is an "outlier" will depend on the underlying distribution and the model used to interpret the data. For example, if the decision is made that the distribution

---

[2] It should be noted that for boxplots, observations marked as "outside values" are simply observations that lie a distance of more than 1.5 times the interquartile range ("IQR") below the 25th percentile or above the 75th percentile. For a normal distribution, the interval that covers [true 25th percentile – 1.5 × IQR, true 75th percentile + 1.5 × IQR] covers 99.3% of the distribution. However, for a skewed distribution this interval covers less of the distribution. As Chambers *et al.* (1983, p. 22) note, "*Outside values are not necessarily outliers* . . . but any outliers will almost certainly appear as outside values." (Emphasis added.)

is other than the normal distribution, then it does not make sense to use Dixon's or Rosner's test on the raw data. USEPA (2009b, p. 6-35) states:

> "A statistical outlier is defined as a value originating from a different statistical population than the rest of the sample. Outliers or observations not derived from the same population as the rest of the sample violate the basic statistical assumption of identically-distributed measurements. If an outlier is suspected, an initial helpful step is to construct a probability plot of the ordered sample data versus the standardized normal distribution (Chapter 12). A probability plot is designed to judge whether the sample data are consistent with a normal population model. ***If the data can be normalized, a probability plot of the transformed observations should also be constructed.*** Neither is a formal test, but can still provide important visual evidence as to whether the suspected outlier(s) should be further evaluated.

> "***Formal testing for outliers should be done only if an observation seems particularly high compared to the rest of the sample***. The data can be evaluated with either Dixon's or Rosner's tests (Chapter 12). These outlier tests assume that the rest of the data except for the suspect observation(s), are normally distributed (Barnett and Lewis, 1994). It is recommended that tests also be conducted on transformed data, if the original data indicates one or more potential outliers. Lognormal and other skewed distributions can exhibit apparently elevated values in the original concentration domain, but still be statistically indistinguishable when normalized via a transformation. ***If the latter is the case, the outlier should be retained and the data set treated as fitting the transformed distribution***."

(Emphasis added).

Additionally, EPA guidance documents point out that the Rosner's test assumes the number of potential outliers is specified in advance and that the data must be examined to get an idea of how many potential outliers there might be. USEPA (2013, p. 189) states:

> "Outliers are not known in advance. ProUCL has normal Q-Q plots which can be used to get an idea about the number of outliers (or mixture populations) potentially present in a data set. This can help a user to determine the suspected number of outliers needed to perform the Rosner test."

Similarly, USEPA (2009a, p. 227) states:

> "In order to use this test, the user has to obtain an initial guess about the number of outliers that may be present in the data set. This can be done by using graphical displays such as a Q-Q plot. On this graphical Q-Q plot, higher observations that are well separated from the rest of the data may be considered to be potential or suspected outliers.

EPA did not provide any evidence that it used the graphical presentations to aid in identifying the number of potential outliers. Instead, contrary to guidance, it ran tests for outliers in an automated fashion, always assuming there are up to 10 outliers. As noted above, EPA

guidance documents state the data should be evaluated to get an idea of how many potential outliers there might be. USEPA (2009b, p. 5-6) points out there are problems with automated outlier removal:

> ***But strategies that involve automated evaluation and removal of outliers may unwittingly eliminate the evidence of real and important changes to background conditions***. An example of this phenomenon may have occurred during the 1970s in some early ozone depletion measurements over Antarctica (http://www.nas.nasa.gov/About/Education/Ozone/history.html). Automated computer routines for outlier detection apparently removed several measurements indicating a sharp reduction in ozone concentrations, and thus prevented identification of an enlarging ozone hole by many years. Later review of the raw observations revealed that these automated routines had statistically classified measurements as outliers, which were more extreme than most of the data from that time period.

(Emphasis added).

Figure 2 helps demonstrate the potential errors resulting from (1) erroneously assuming normal distribution and (2) *a priori* assuming the number of outliers without proper analysis. The figure displays the results of a simulation study showing the behavior of Rosner's test in the case of various underlying distributions. For each kind of distribution, 50 observations from that distribution were generated, and Rosner's test was applied to these data setting $k = 10$ potential outliers and assuming a Type I error level of 5%. This procedure was repeated 1,000 times. Figure 1 shows the distribution of the number of outliers indicated by Rosner's test. For data generated from a normal (Gaussian) distribution, most of the time no observations were determined to be "outliers," and the proportion of times at least one observation was labeled as an outlier was 0.04, with a 95% confidence interval ("CI") of [0.03, 0.06], which encompasses the assumed Type I Error level. For data generated from a gamma distribution with mean = 10 and coefficient of variation ("CV") = 1, the proportion of times at least one observation was labeled as an outlier was 0.76, with a 95% CI of [0.73, 0.78]. Similarly, for a lognormal distribution with mean = 10 and CV = 1, the proportion of times at least one observation was labeled as an outlier was 0.90, with a 95% CI of [0.88, 0.92]. In cases when the true underlying distribution is not normal, Rosner's test can incorrectly identify data as "outliers" because of a grossly inflated Type I error level.

3.   EPA Abused Its Discretion by Discarding Data Without Investigating Whether Evidence Justified Removal of the Data

USEPA (2009b, pp. 5-5 to 5-6) states:

"Outliers or observations not derived from the same population as the rest of the sample violate the basic statistical assumption of identically-distributed measurements. ***The Unified Guidance recommends that testing of outliers be performed on background data, but they generally not be removed unless some basis for a likely error or discrepancy can be identified***. Such possible errors or discrepancies could include data recording errors, unusual sampling and

laboratory procedures or conditions, inconsistent sample turbidity, and values significantly outside the historical ranges of background data. Management of potential outliers carries both positive and negative risks, which should be carefully understood. . . .

"***Ideally, removal of one or more statistically identified outliers should be based on other technical information or knowledge which can support that decision***."

(Emphasis added).

USEPA (2013, p. 189) states:

"***The outlying observations should be investigated separately to determine the reasons for their occurrences (e.g., errors or contaminated locations)***. It is suggested to compute the statistics with and without the outliers, and compare the potential impact of outliers on the decision making processes."

(Emphasis added).

Contrary to this guidance, EPA is computing BTVs for upriver bedded sediment based on a data set that excludes specific data points that the EPA has improperly deemed to be "outliers" based on the misapplication of Rosner's or Dixon's test and that EPA chose to remove based on its "visual observations." Again, data cannot be discarded based simply on a statistical test, especially when, as here, it was an error to use Rosner's or Dixon's test in the first place. That is because those tests assume a normal distribution of data, which is not necessarily the case here as discussed above. EPA must investigate whether or not there is any evidence to justify discarding these observations. *See* USEPA (2009b, pp. 5-5, 5-6).

EPA has not identified any likely errors or discrepancies that provide a basis for excluding the suspected outliers. The only evidence provided regarding a suspected source is anecdotal information about Portland Shipbuilding Company ("PSC") at river mile 16, but EPA did not locate a comprehensive history of PSC's activities that would support a conclusion that it is a source of the excluded concentrations. EPA then ignored the fact that deleted "outliers" were also associated with RM 17, RM 18.5, RM 23, RM 24, RM 27, and RM 28.5. EPA also ignored more likely explanations for the higher concentrations, including the PCB-binding capacity of the sediments (see Section 5 below) in that specific section of the river. Thus, in addition to the issue of "outliers" being identified by statistically unsound methods, EPA's decision to remove them is not supported by any other lines of evidence. Further, even if some of the concentrations are due to contaminants from the PSC, there are no plans for further investigation or remediation and therefore no reason to remove these observations since upstream bedded sediments with elevated concentrations may be transported downstream to the Site. Thus, it is important for the reference area data to represent the ***total*** reference area population.

4.   EPA Did Not Use the Correct Procedures for Data Containing Nondetect Values

ProUCL 5.0.00 Technical Guide (USEPA 2013, p. 27) states:

"Based upon the results of the report by Singh, Maichle, and Lee (EPA, 2006), it is recommended to avoid the use of the DL/2 method to perform a GOF test, and to compute the summary statistics and various other limits (e.g., UCL, UPL, UTLs) often used to estimate the EPC terms and BTVs."

This guidance is consistent with the standard statistical approach. EPA, however, ignored it in favor of other recommendations in USEPA 2009a and USEPA 2013 that deviate from the standard approach and lead to incorrect statistical results. The following discusses the use of Q-Q plots, GOF tests and outlier tests for data containing nondetects and notes where EPA both in interpretation of guidance and in consideration of the data at issue in this dispute incorrectly deviated from standard approaches.

a.   Q-Q Plots for Data that Contain Nondetects

If a dataset contains nondetect values, specific statistical methods are available for computing the correct plotting positions (e.g., USEPA, 2009b; Helsel, 2012; Millard, 2013). However, USEPA (2009a, p. 84) and USEPA (2013, p. 33) incorrectly recommend: (1) excluding all nondetects and then constructing the Q-Q plot, (2) setting nondetects to half the detection limit and then constructing the Q-Q plot, or (3) setting nondetects to the detection limit and then constructing the Q-Q plot. Method 1 is simply throwing away data that one does not know what to do with and is patently wrong. Methods 2 and 3 are incorrect, as noted subsequently by EPA in USEPA (2009b, p. 12-2):

"A related difficulty occurs when sample data includes censored or non-detect values. If simple substitution is used to estimate a value for each non-detect prior to plotting, the resulting probability plot may appear non-linear simply because the censored observations were not properly handled. In this case, a censored probability plot . . . should be constructed instead of an uncensored, complete sample plot . . . ."

Figures 2.1 – 2.33 show correct normal (Gaussian) Q-Q plots (see USEPA, 2009b; Helsel, 2012; Millard, 2013), for all Indicator Chemicals considered in the final background dataset with adequate sample sizes. These plots were created with the R package *EnvStats* (Millard, 2013). Each page contains four separate Q-Q plots: based on the raw data, based on the square-root of the observations, based on the cube-root of the observations, and based on the natural logarithm of the observations. One can create gamma Q-Q plots as well; however, a cube-root transformation is often recommended to attempt to make data from a gamma distribution appear normal (Kulkarni and Powar, 2010). As an example of the conclusions one can draw from these Q-Q plots, Figure 2.29 shows that Total PCBs Aroclors are adequately modeled by a lognormal distribution.

b.   Goodness-of-Fit Tests for Data that Contain Nondetects

Additionally, if a dataset contains nondetect values, specific correct statistical methods also are available for computing a GOF test for normality (Royston, 1993; Millard, 2013). Again, both USEPA (2013, pp. 10, 111) and USEPA (2009a, p. 110) deviate from these standard approaches and recommend: (1) excluding all nondetects and then performing the GOF test, (2) setting nondetects to half the detection limit and then proceeding as if the original dataset had

no nondetects, or (3) using regression on order statistics ("ROS") methods to impute the values for the nondetects and then proceeding as if the original dataset had no nondetects. Similar to Q-Q plot issue, Method 1 is simply throwing away data that one does not know how to handle. Method 2 is not recommended by USEPA (2013, p.27) as noted above. Method 3 ignores the uncertainty associated with imputing values for the nondetects. See Figures 2.1 – 2.33 for examples of current methods. These include p-values from GOF tests; for analytes with nondetects, the method of Royston (1993) was used. These data are also summarized in the attached Table 1. As an example of the conclusions one can draw from these GOF tests, Figure 2.29 shows that Total PCBs Aroclors are adequately modeled by a lognormal distribution.

c.   Outlier Tests for Data that Contain Nondetects

Although correct GOF tests exist for data with nondetects as explained in the previous section, currently no commonly used standard tests for outliers are available when the data contain nondetect values, although methods do exist. For example, Nardi and Schemper (1999) developed an algorithm to identify outlying observations based on Cox linear regression for censored data, and Eo *et al.* (2014) proposed a similar test based on quantile regression.

In order to test for outliers on datasets that contain nondetects, USEPA (2013, pp. 191-192) again recommends: (1) excluding all nondetects and then performing the test, (2) setting nondetects to half the detection limit and then performing the test, or (3) setting nondetects to the detection limit and then performing the test. USEPA (2009a, p. 223) recommends Methods 1 or 2. As noted previously, Method 1 simply throws away data, and Methods 2 and 3 are statistically incorrect. In general, applying a standard outlier test developed for data with no censored values to data with nondetects, where the nondetects are set to half the detection limit or the detection limit, will produce questionable results.

As an example, Figures 3.1 – 3.3 display the results of a simulation study showing the behavior of Rosner's test in the case of various underlying distributions and censoring. Detection limits were chosen to be at the 15th, 30th, 45th, or 60th percentile of the distribution, resulting in singly censored data. For each kind of distribution and each censoring level, 50 observations from that distribution were generated, then all observations less than the specified detection limit were treated as "nondetects." Nondetects were set to half the detection limit as prescribed by USEPA (2013), and then Rosner's test was applied to these data setting $k = 10$ potential outliers and assuming a Type I error level of 5%. This procedure was repeated 1,000 times. Figure 3.1 shows the results based on generating data from a normal (Gaussian) distribution. As in Figure 1, the distribution of the number of outliers indicated by Rosner's test is shown. With a detection limit set to the 15th percentile, resulting in 15% censoring on average, the proportion of times at least one observation was labeled as an outlier was 0.84, with a 95% CI of [0.82, 0.87], clearly showing that the assumed Type I error is nowhere near the assumed 5% for this scenario. For the other censoring levels, the Type I error is not as grossly inflated, probably because as more and more observations become censored and set to the same value, differences between any observation and the mean decreases; this would require further investigation to verify. Figures 3.2 and 3.3 show results based on the same gamma and lognormal distributions used in Figure 1. Because for these distributions the Type I error is already hugely over-inflated (see Figure 1), the incorrect handling of censored observations does not affect the results as much as in the case of an underlying normal distribution. Note that these

14

results are for the simple case of single censoring. Other simulations could be performed using more complicated censoring.

5.   Physical Characteristics of the "Outlier" Samples

As discussed above, USEPA (2002a, p. 2-2) states that the locations and characteristics of the ***sampling stations*** used in the background area ***should reflect*** the physical, chemical, geological, and biological characteristics of ***the area itself***. Furthermore, the characteristics of the sampling stations in the background area should be comparable to the characteristics (not related to activities associated with the Study Area) of the sampling stations that will be used to characterize the Study Area.

It is clear from EPA's 2002 Role of Background in the CERCLA Cleanup Program that EPA expected some sites to have elevated anthropogenic background concentrations:

"Background refers to constituents or locations that are not influenced by the releases from a site, and is usually described as naturally occurring or anthropogenic (EPA, 1989; EPA, 1995a):

1) Anthropogenic – natural and human-made substances present in the environment as a result of human activities (not specifically related to the CERCLA release in question); and,

2) Naturally occurring – substances present in the environment in forms that have not been influenced by human activity."

This document goes on to state:

> "In cases where background levels are high or present health risks, this information may be important to the public. Background information is important to risk managers because the CERCLA program, generally, does not clean up to concentrations below natural or ***anthropogenic background*** levels." (Emphasis added)

Therefore, it should not be surprising to have upstream bedded sediments exhibiting elevated anthropogenic concentrations. In fact, many of the reference area samples exhibiting the highest contaminant concentrations and removed as outliers based on EPA's flawed statistical approach were in relatively quiescent or sheltered areas of the reference area and had percent fines and organic carbon content that best approached the sediments in the majority of the Study Area. For example, the average organic carbon content of all surface sediment samples from the Study Area equals 1.79%, this compares to an average carbon content of 1.11% for all reference area samples. But the average organic carbon content of background samples identified and removed as outliers equals 1.66%. The reference area samples with lower contaminant concentrations were obtained in areas known to be higher energy than most of the Study Area and exhibited lower percent fines and organic material. As a result, EPA's background statistical methodology identified and removed as "outliers" data that actually best represented background conditions for the Study Area. The general impacts of percent fines and organic carbon on contaminant concentrations of sediment were thoroughly discussed in the Draft RI Report. EPA

Region 10 staff chose to ignore the importance of these factors in their revision of the RI report and development of background values, despite overwhelming evidence of their importance in the scientific literature, acknowledgement in EPA's 2005 sediment remediation guidance (EPA 2005), and site-specific analysis provided in the Draft RI. The analysis presented in the Draft RI was conducted according to specific discussions between EPA and the LWG in developing the analysis process for the RI and preparing the text in the Draft RI.

Conclusion

The full background data set with consideration of organic carbon correction is the most appropriate data set for determining BTVs. Thus, the LWG requests that the full data set with consideration of organic carbon correction be retained as the selected set of background values and applied in the FS. These values are shown in the "all data" columns of Table 7.3-1b (and the related Appendix H Table H-2b) of the RI Section 7 revision agreed to by EPA and the LWG on December 12, 2013. This approach is consistent with EPA guidance and the NCP, it is based on sound science, and it is supported by the other strong lines of evidence presented in the RI.

As noted above, Dr. Steve Millard has assisted the LWG in analyzing EPA's statistical approach, and he is available to meet with you to answer any questions. A summary of his credentials is enclosed. Additionally, if requested by EPA, the LWG will provide further analysis and support for each of the above objections.

Sincerely,

The Lower Willamette Group

cc: Lori Cohen, EPA Region 10 Associate Director, Office of Environmental Cleanup
        Deborah Yamamoto, EPA Region 10 Office of Environmental Cleanup
        Kristine Koch, EPA Region 10 Office of Environmental Cleanup
        Sean Sheldrake, EPA Region 10 Office of Environmental Cleanup
        Lori Cora, EPA Region 10 Assistant Regional Counsel
        Jim Woolford, EPA Headquarters
        Barry Nussbaum, EPA Headquarters
        Confederated Tribes and Bands of the Yakama Nation (via EPA Shared Server)
        Confederated Tribes of the Grand Ronde Community of Oregon (via EPA Shared Server)
        Confederated Tribes of Siletz Indians of Oregon (via EPA Shared Server)
        Confederated Tribes of the Umatilla Indian Reservation (via EPA Shared Server)
        Confederated Tribes of the Warm Springs Reservation of Oregon (via EPA Shared Server)
        Nez Perce Tribe (via EPA Shared Server)
        Oregon Department of Fish & Wildlife (via EPA Shared Server)
        United States Fish & Wildlife (via EPA Shared Server)
        Oregon Department of Environmental Quality (via EPA Shared Server)
        LWG Legal
        LWG Repository

References

Box, G.E.P., and N.R. Draper. (1987). *Empirical Model-Building and Response Surfaces*. John Wiley & Sons, New York.

Chambers, J.M., W.S. Cleveland, B. Kleiner, and P.A. Tukey. (1983). *Graphical Methods for Data Analysis*. Duxbury Press, Boston, MA.

Eo, S-H., S.M. Hong, and H. Cho. (2014). Identification of Outlying Observations with Quantile Regression for Censored Data. Submitted.

Helsel, D.R. (2012). *Statistics for Censored Environmental Data Using Minitab and R*. Second Edition. John Wiley & Sons, Hoboken, New Jersey.

Kulkarni, H.V., and S.K. Powar. (2010). A New Method for Interval Estimation of the Mean of the Gamma Distribution. *Lifetime Data Analysis*, **16**, 431–447.

Millard, S.P. (2013). *EnvStats: An R Package for Environmental Statistics*. Springer, New York.

Millard, S.P., and N.K. Neerchal. (2001). *Environmental Statistics with S-PLUS*. CRC Press, Boca Raton, Florida.

Nardi A., and M. Schemper. (1999). New Residuals for Cox Regression and their Application to Outlier Screening. *Biometrics* **55**(2), 523–529.

Royston, P. (1993). A Toolkit for Testing for Non-Normality in Complete and Censored Samples. *The Statistician* **42**, 37-43.

USEPA. (2002a). *Guidance for Comparing Background and Chemical Concentrations in Soil for CERCLA Sites*. EPA 540-R-01-003, September, 2002. Office of Emergency and Remedial Response, U.S. Environmental Protection Agency, Washington, D.C.

USEPA. (2002b). *Role of Background in the CERCLA Cleanup Program*. OSWER 9285.6-07P, April, 2002. Office of Solid Waste and Emergency Response, Office of Emergency and Remedial Response, U.S. Environmental Protection Agency, Washington, D.C.

USEPA. (2005). *Contaminated Sediment Remediation Guidance for Hazardous Waste Sites*. OSWER 9355.0-85, December, 2005. Office of Solid Waste and Emergency Response, Office of Emergency and Remedial Response, U.S. Environmental Protection Agency, Washington, D.C.

USEPA. (2006a). *Data Quality Assessment: A Reviewer's Guide*. EPA QA/G-9R. EPA/240/B-06/002, February 2006. Office of Environmental Information, U.S. Environmental Protection Agency, Washington, D.C.

USEPA. (2006b). *Data Quality Assessment: Statistical Methods for Practitioners*. EPA QA/G-9S. EPA/240/B-06/003, February 2006. Office of Environmental Information, U.S. Environmental Protection Agency, Washington, D.C.

USEPA. (2009a). *Scout 2008 Version 1.0 User Guide (Second Edition, December 2008).* EPA/600/R-08/038, February 2009. Office of Research and Development, National Exposure Research Laboratory, Environmental Sciences Division, Technology Support Center, Characterization and Monitoring Branch, U.S. Environmental Protection Agency, Las Vegas, Nevada.

USEPA. (2009b). *Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities: Unified Guidance.* EPA 530-R-09-007, March 2009. Office of Resource Conservation and Recovery, Program Implementation and Information Division, U.S. Environmental Protection Agency, Washington, D.C.

USEPA. (2013). *ProUCL Version 5.0.00 Technical Guide*. EPA/600/R-07/041, September 2013. Office of Research and Development, U.S. Environmental Protection Agency, Washington, D.C.

van Belle, G. (2008). *Statistical Rules of Thumb*. Second Edition. John Wiley & Sons, New York.

Table

Table 1. Goodness-of-Fit Test Results for Indicator Chemicals.

Analyte	Number Stations Sampled	Percent Stations Sampled[1]	Number Nondetect	Percent Nondetect[2]	GOF Test[3]	P-value			
						Raw Data	Square Root	Cube Root	Log
Aldrin	48	68	42	88	S-F	1	1	1	0.98
Arsenic	71	100	0	0	S-W	< 0.001	< 0.001	< 0.001	0.01
Benzo(a)pyrene	52	73	12	23	S-F	< 0.001	< 0.001	0.003	0.001
Bis(2-ethylhexyl) phthalate	48	68	4	8	S-F	< 0.001	< 0.001	< 0.001	0.03
Butylbenzyl phthalate	48	68	33	69	S-F	0.02	0.31	0.53	0.65
Cadmium	67	94	9	13	S-F	0.03	0.11	0.06	0.005
Chromium	65	92	0	0	S-W	0.36	0.86	0.89	0.76
Copper	67	94	0	0	S-W	0.03	0.22	0.26	0.20
CPAH	52	73	8	15	S-F	< 0.001	< 0.001	0.008	0.002
Dieldrin	48	68	41	85	S-F				

						0.98	0.95	0.91	0.78
Diesel Range Hydrocarbons	28	39	0	0	S-W	0.98	0.20	0.02	< 0.001
Dioxin TEQ - Mammals 2006	52	73	0	0	S-W	< 0.001	< 0.001	< 0.001	0.05
gamma-Hexachlorocyclohexane	48	68	47	98	S-F	NA	NA	NA	NA
Hexachlorobenzene	48	68	24	50	S-F	0.002	0.03	0.08	0.34
Lead	67	94	0	0	S-W	< 0.001	< 0.001	< 0.001	< 0.001
Mercury	61	86	9	15	S-F	0.65	0.88	0.65	0.09
Naphthalene	52	73	33	63	S-F	0.99	0.99	0.96	0.70
Nickel	67	94	0	0	S-W	0.61	0.21	0.11	0.02
PCB TEQ - Mammals 2006	33	46	1	3	S-F	< 0.001	< 0.001	< 0.001	< 0.001
Pentachlorophenol	52	73	50	96	S-F	1	1	1	1
Phenanthrene	52	73	22	42	S-F	< 0.001	< 0.001	< 0.001	< 0.001
Residual Range Hydrocarbons	28	39	0	0	S-W	0.30	0.02	0.002	< 0.001

Analyte	Number Stations Sampled	Percent Stations Sampled[1]	Number Nondetect	Percent Nondetect[2]	Test[3]				
Total Chlordane (calc'd)	48	68	15	31	S-F	0.04	0.04	0.01	< 0.001
Total DDTs (calc'd)	48	68	1	2	S-F	< 0.001	0.04	0.02	< 0.001
Total HPAHs (calc'd)	52	73	8	15	S-F	< 0.001	0.005	0.02	< 0.001
Total LPAHs (calc'd)	52	73	12	23	S-F	< 0.001	< 0.001	< 0.001	< 0.001
Total PAHs (calc'd)	52	73	7	13	S-F	< 0.001	0.01	0.03	< 0.001
Total PCB Congeners (calc'd)	33	46	0	0	S-W	< 0.001	< 0.001	< 0.001	0.06
Total PCBs Aroclors (calc'd)	48	68	25	52	S-F	0.009	0.10	0.19	0.49
Total PCDD/F (calc'd)	33	46	0	0	S-W	< 0.001	0.02	0.07	0.20
Total Petroleum Hydrocarbons (calc'd)	28	39	0	0	S-W	0.41	0.03	0.002	< 0.001
Tributyltin ion	3	4	1	33	S-F	NA	NA	NA	NA
Zinc	67	94	0	0	S-W	0.001	0.08	0.19	0.37

[1] Total number of stations sampled in the background area is 71. Percent Stations Sampled = 100 × Number Stations Sampled / 71

[2] Percent Nondetect = 100 × Number Nondetect / Number Stations Sampled

[3] S-W = Shapiro-Wilk Goodness-of-Fit test; S-F = Shapiro-Francia Goodness-of-Fit Test (Royston, 1993).

Figures

Figure 1. Distribution of Number of Outliers According to Rosner's Test for Various Probability Distributions

Distribution of Number of Outliers According to Rosner's Test

k = 10; Assumed Type I Error = 5%; Number of Simulations = 1,000
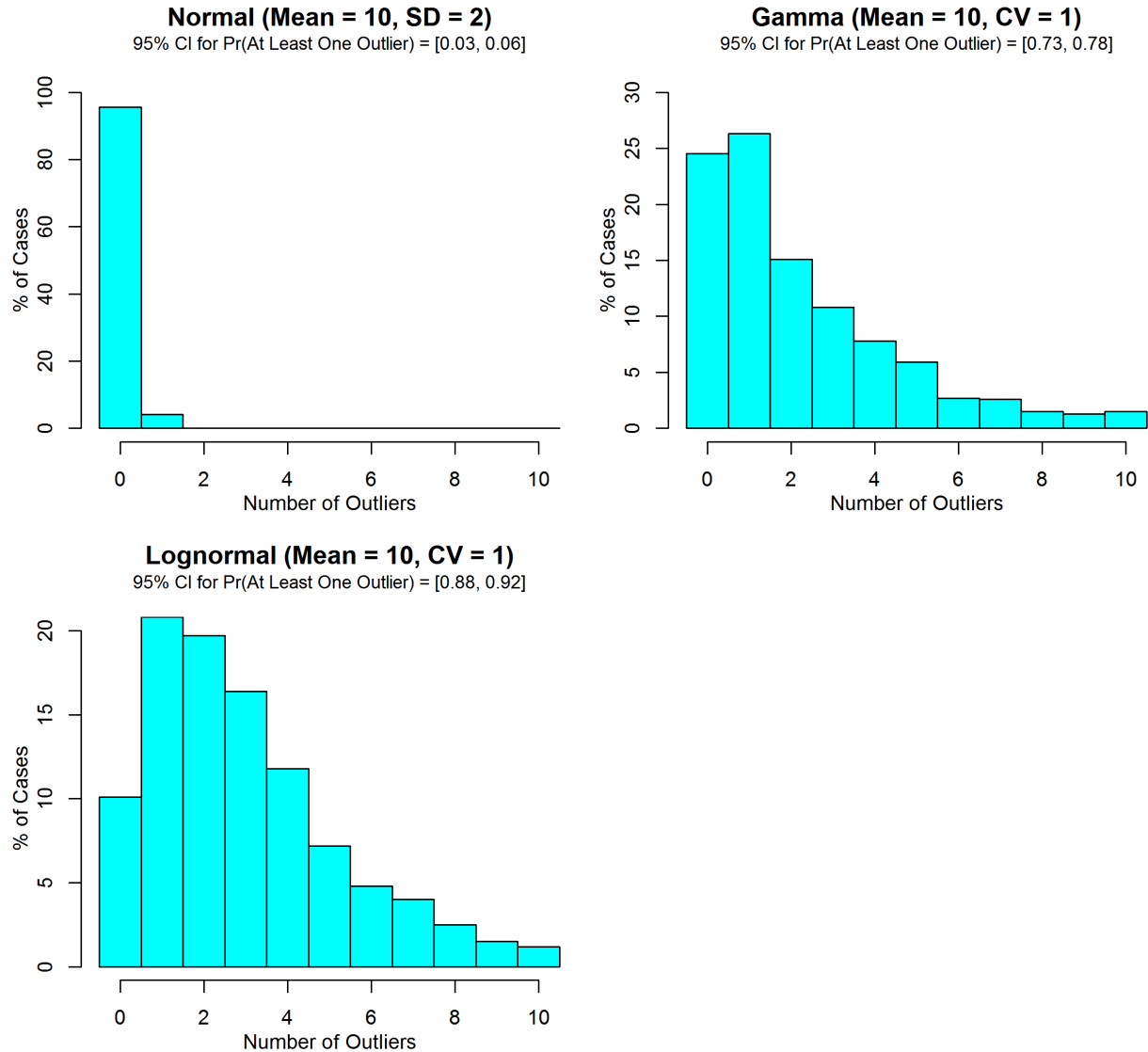
Figure 2: Normal Q-Q Plots

Figure 2.1: Censored Normal Q-Q Plots for Aldrin

Censored Normal Q-Q Plots for Aldrin (ug/kg)
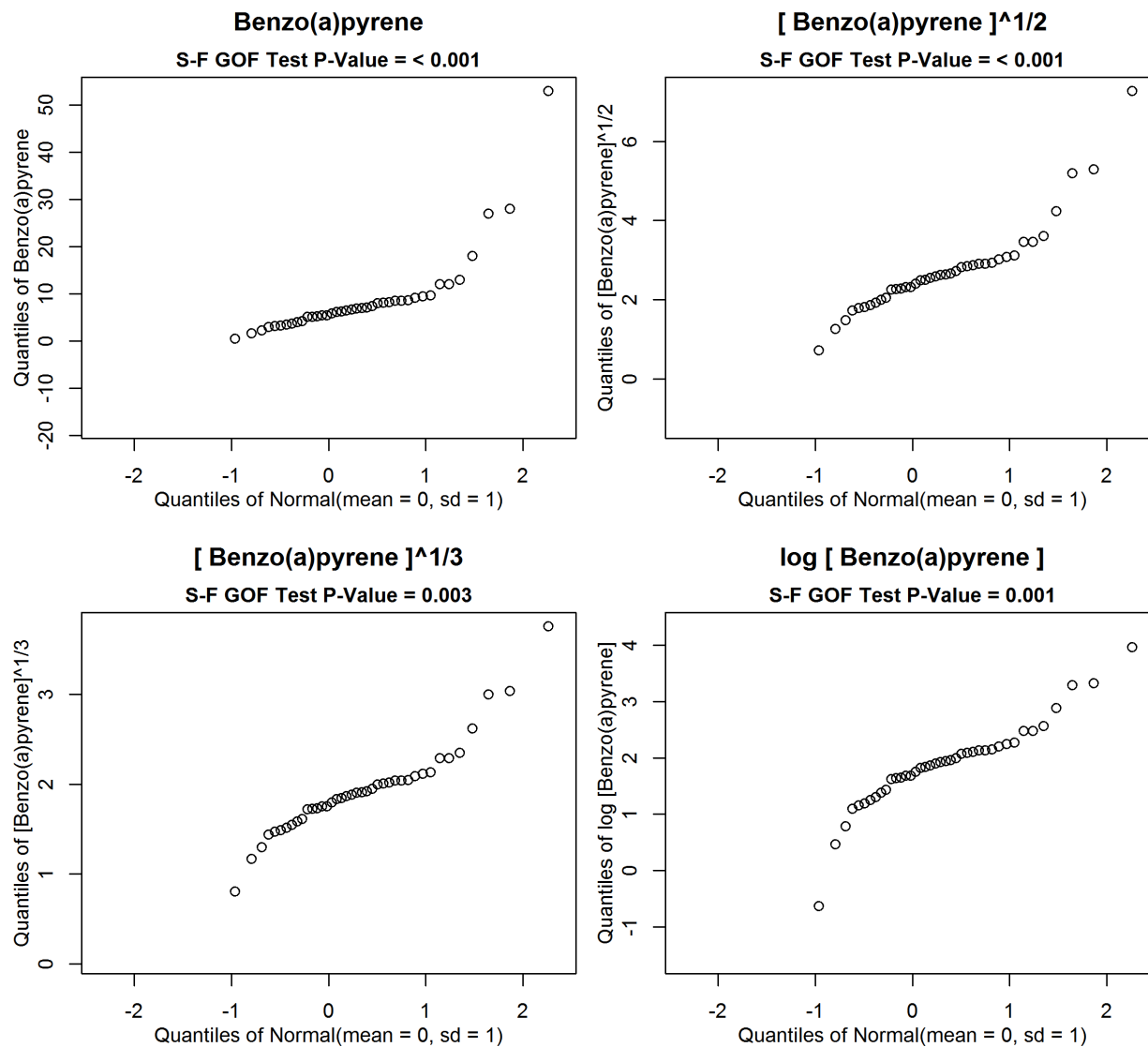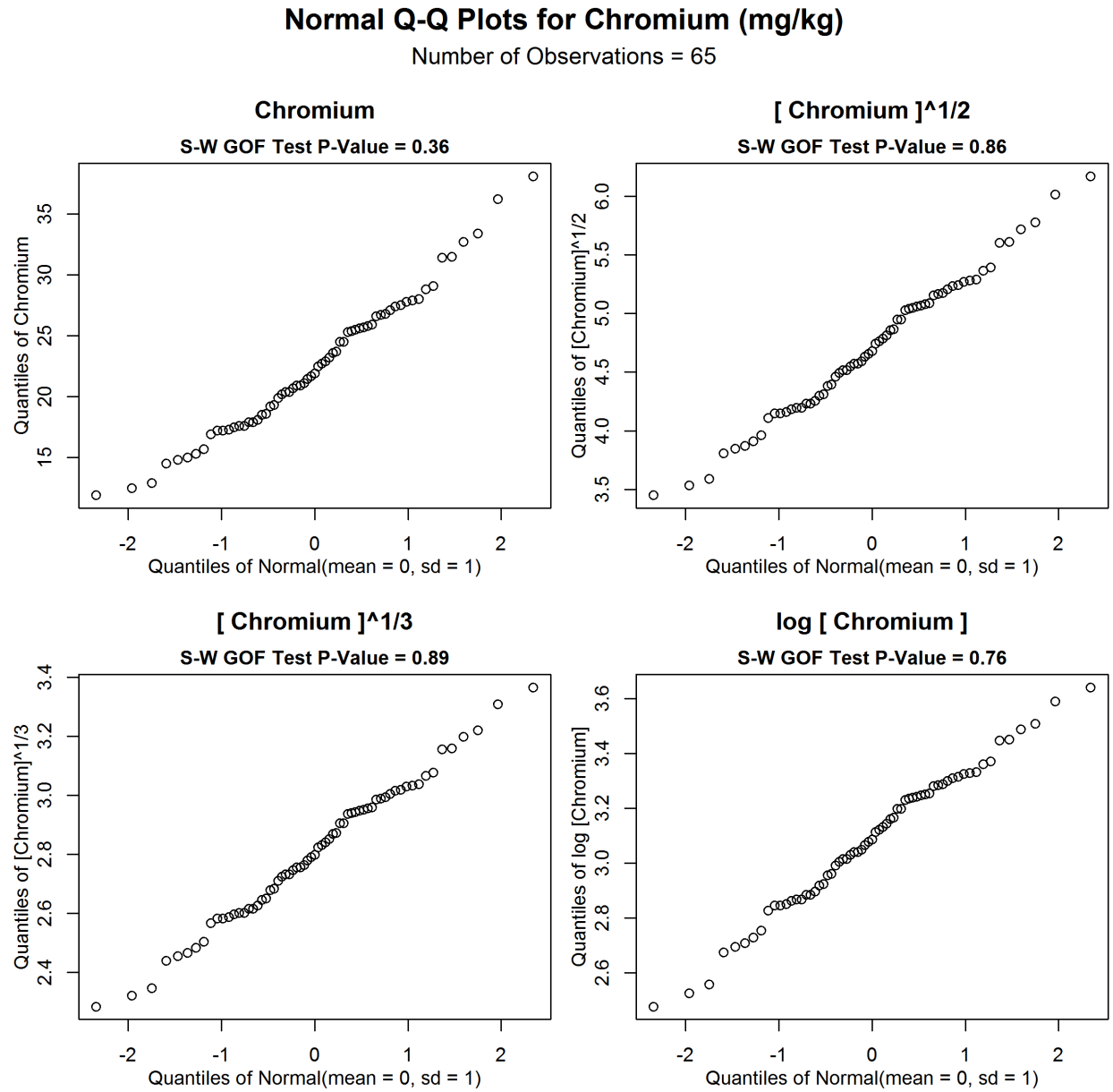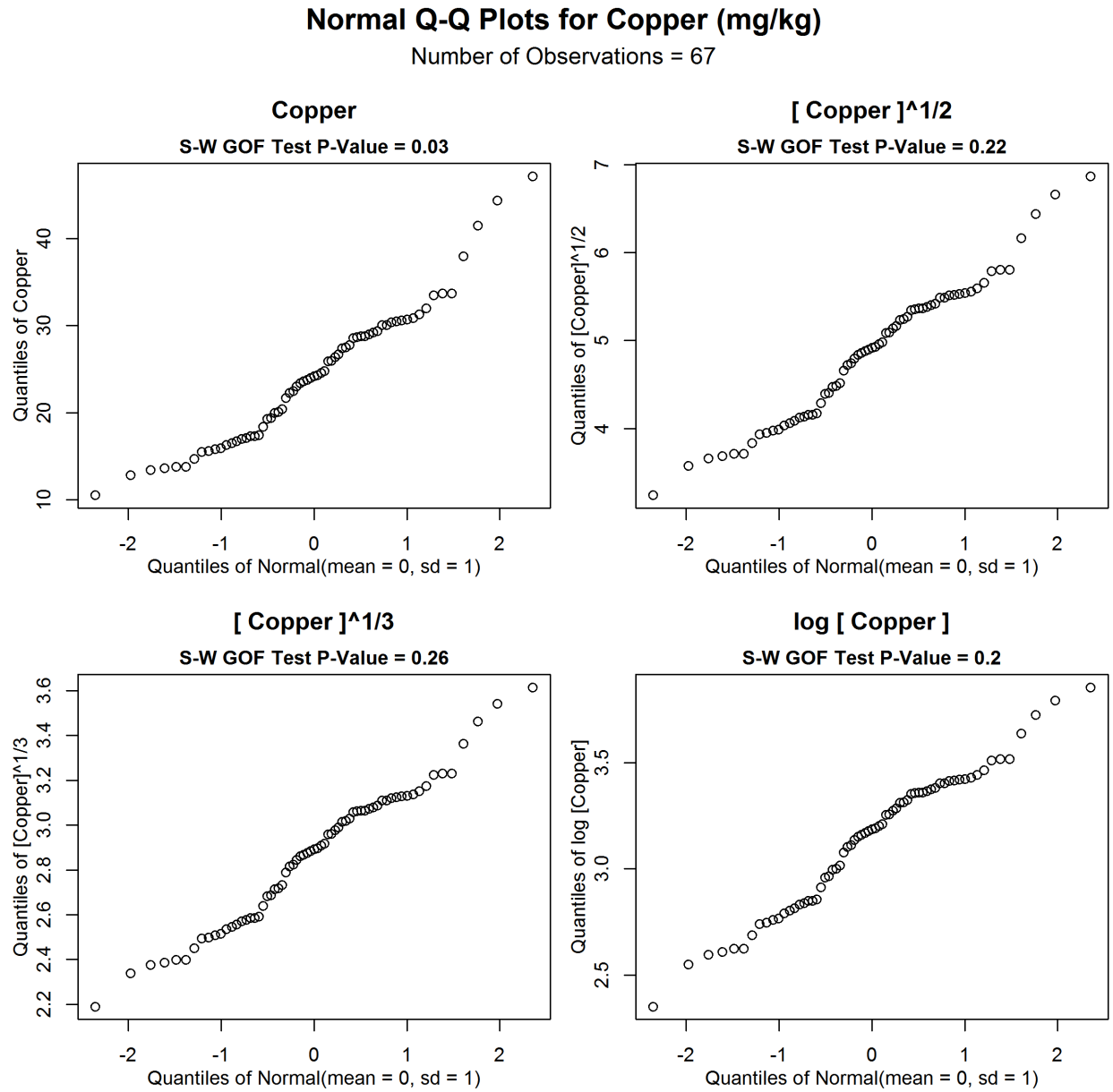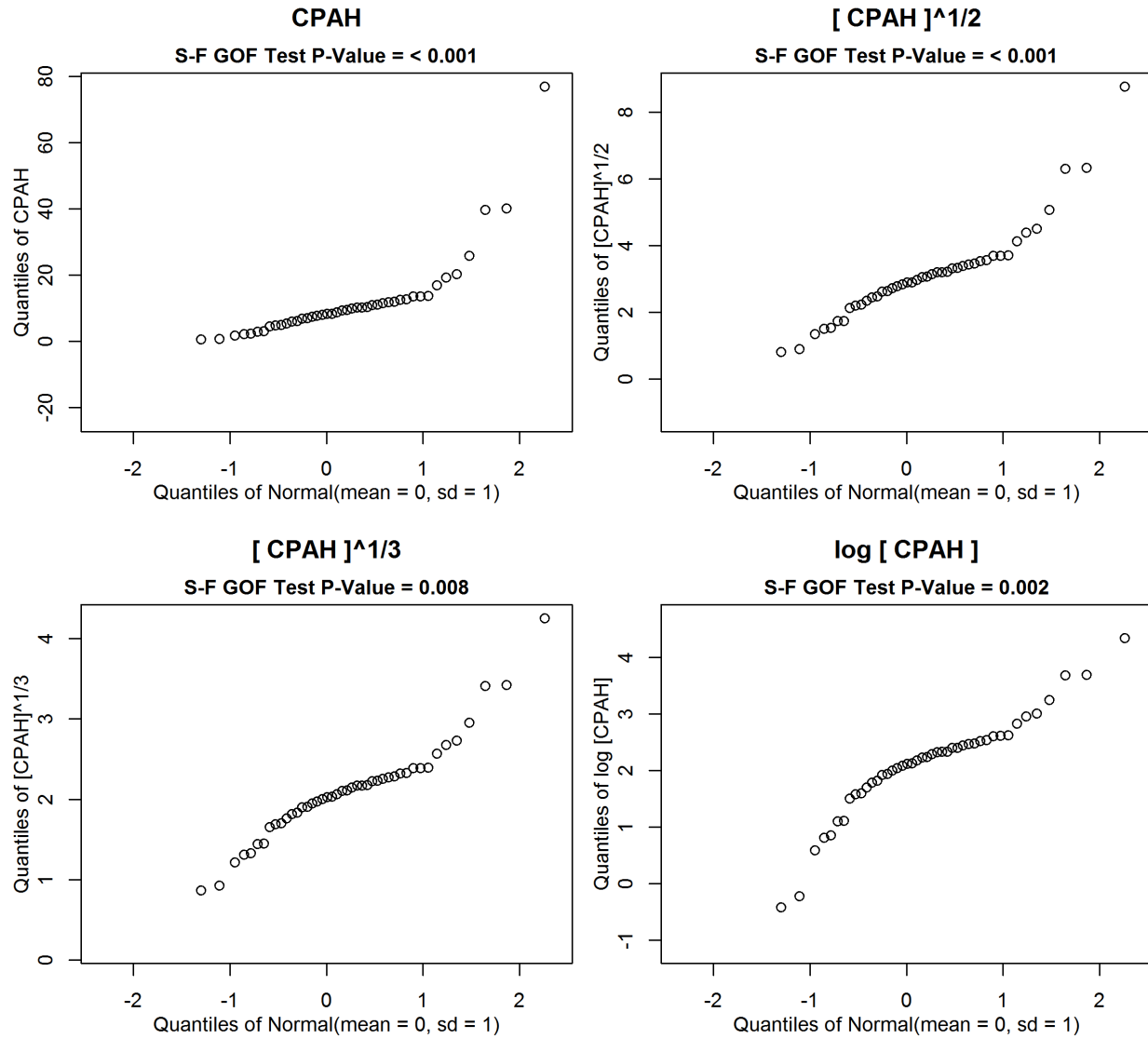
Number of Censored Observations = 42 out of 48

Aldrin — S-F GOF Test P-Value = 1

[Aldrin]^1/2 — S-F GOF Test P-Value = 1

[Aldrin]^1/3 — S-F GOF Test P-Value = 1

log [Aldrin] — S-F GOF Test P-Value = 0.98

23
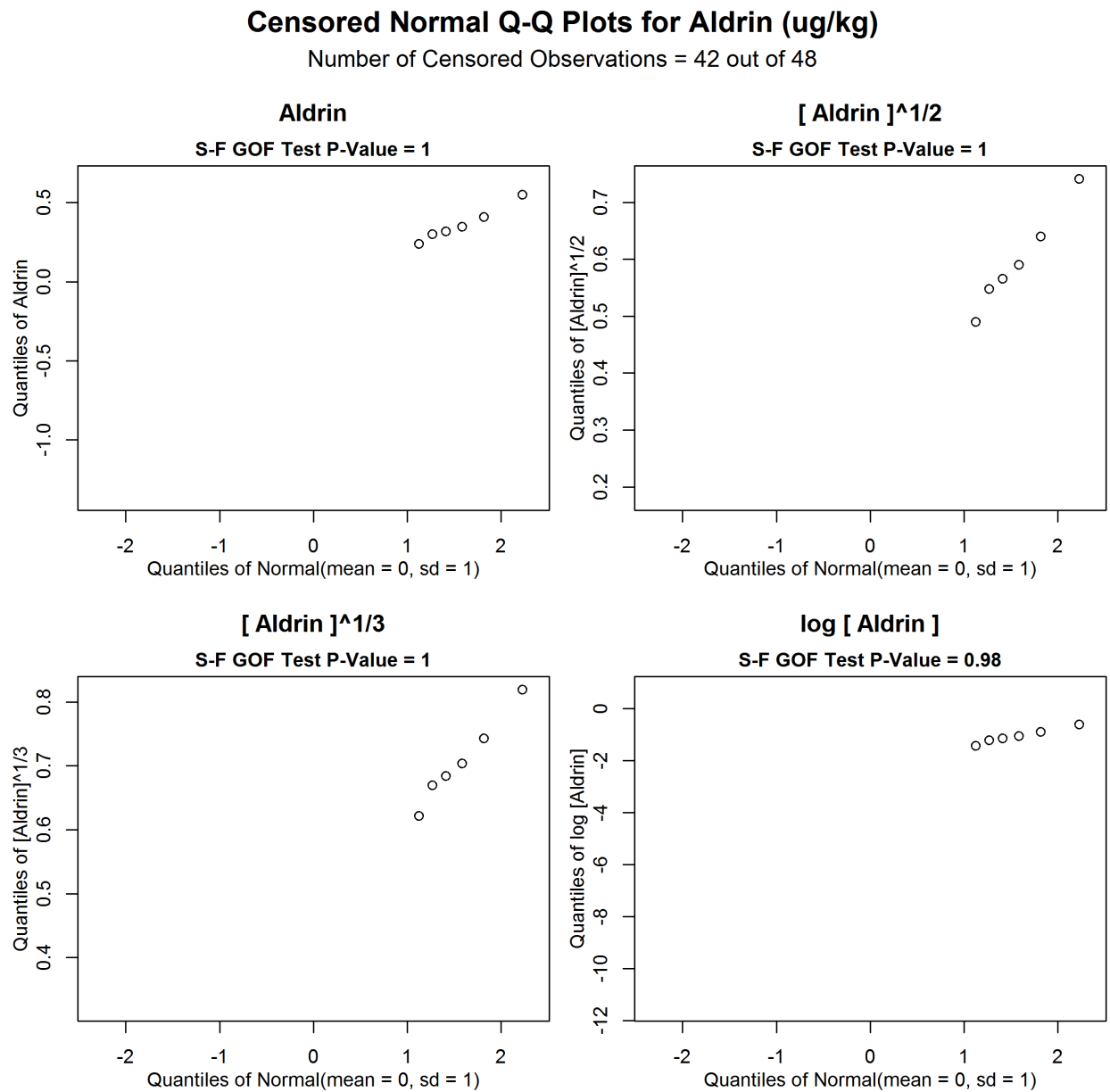
Figure 2.2: Normal Q-Q Plots for Arsenic

Normal Q-Q Plots for Arsenic (mg/kg)

Number of Observations = 71
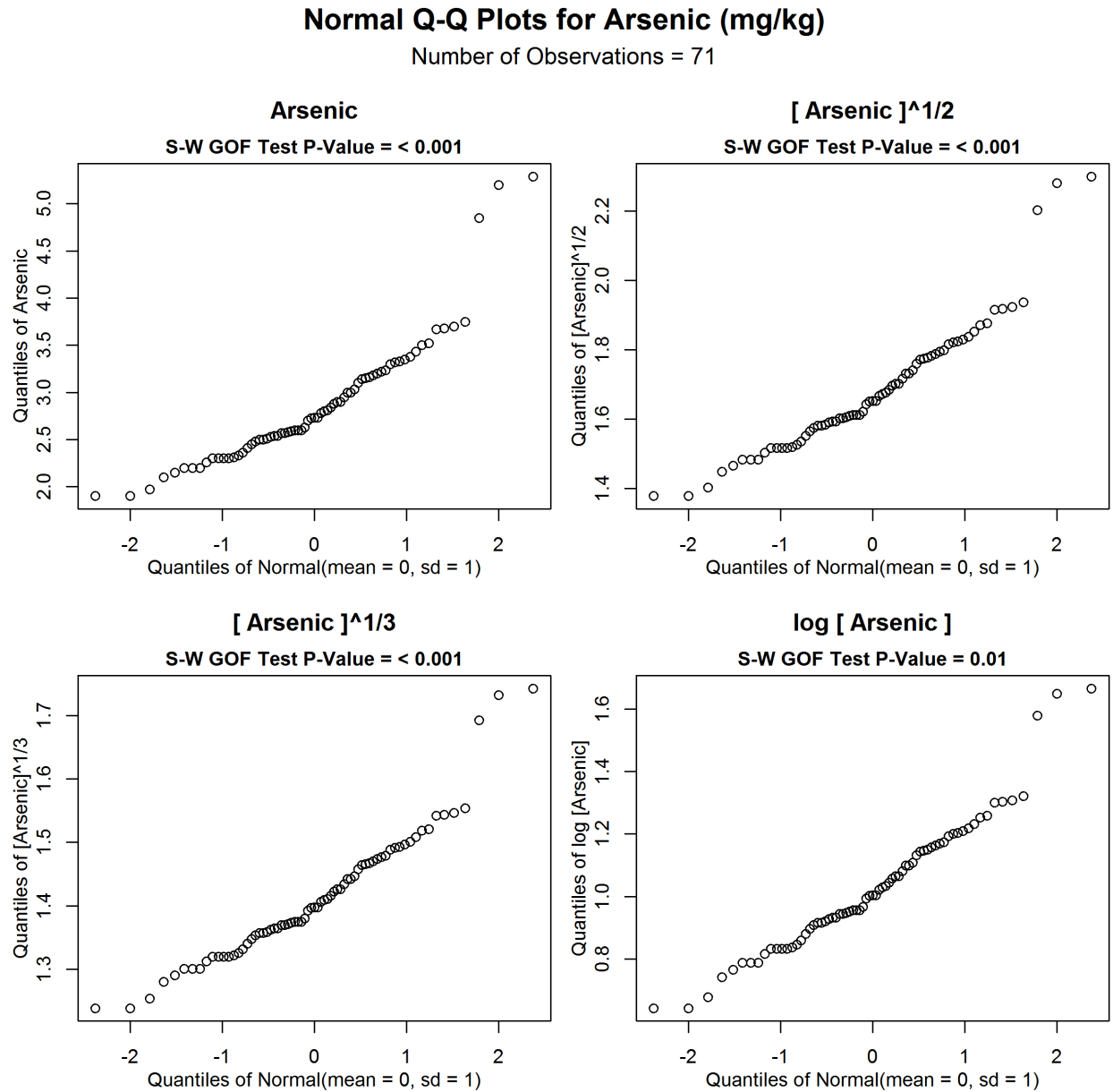
Figure 2.3: Censored Normal Q-Q Plots for Benzo(a)pyrene

Censored Normal Q-Q Plots for Benzo(a)pyrene (ug/kg)

Number of Censored Observations = 12 out of 52

Benzo(a)pyrene — S-F GOF Test P-Value = < 0.001

[Benzo(a)pyrene]^1/2 — S-F GOF Test P-Value = < 0.001

[Benzo(a)pyrene]^1/3 — S-F GOF Test P-Value = 0.003

log [Benzo(a)pyrene] — S-F GOF Test P-Value = 0.001

Figure 2.4: Censored Normal Q-Q Plots for Bis(2-ethylhexyl) phthalate

Censored Normal Q-Q Plots for Bis(2-ethylhexyl) phthalate (ug/kg)

Number of Censored Observations = 4 out of 48
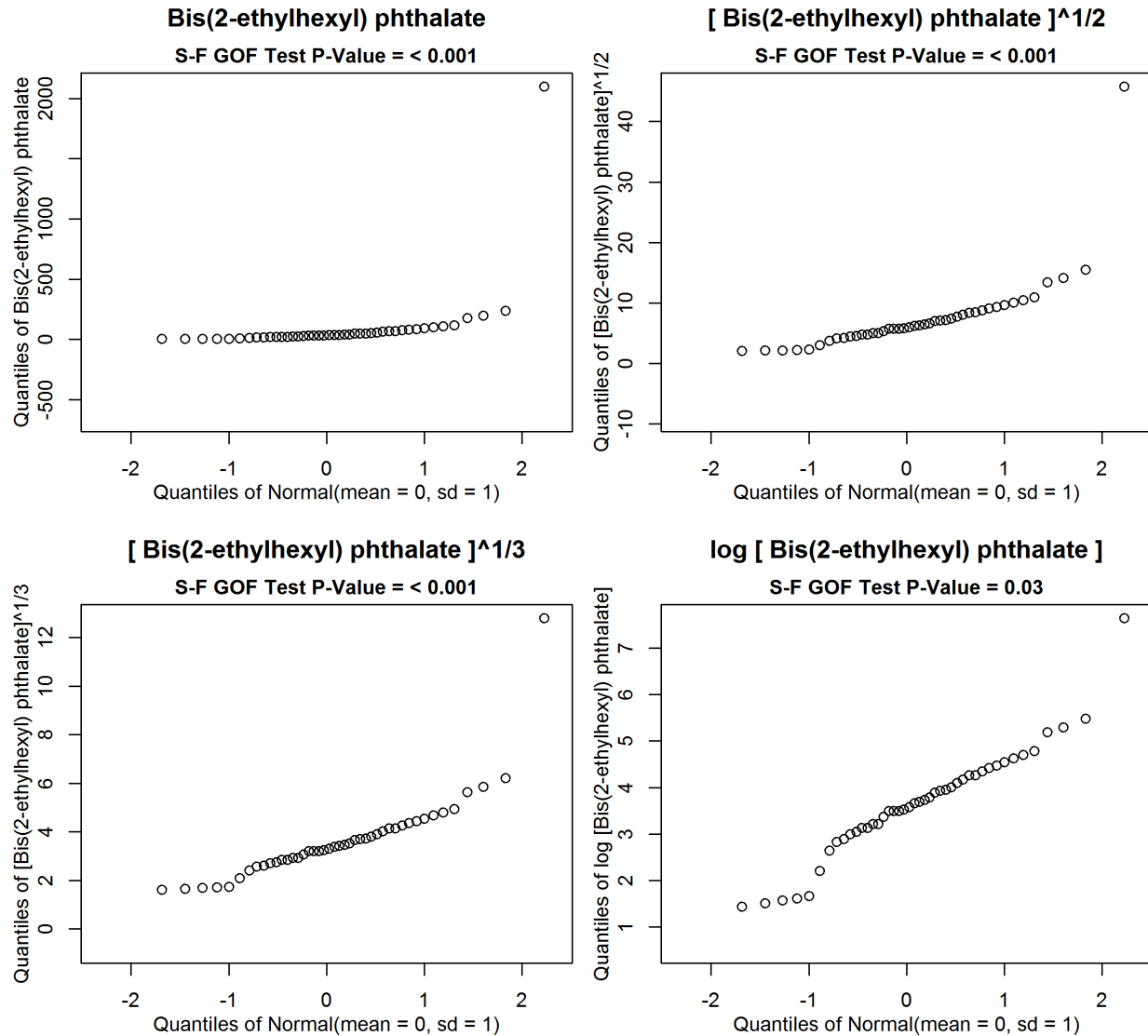
Figure 2.5: Censored Normal Q-Q Plots for Butylbenzyl phthalate

Censored Normal Q-Q Plots for Butylbenzyl phthalate (ug/kg)

Number of Censored Observations = 33 out of 48
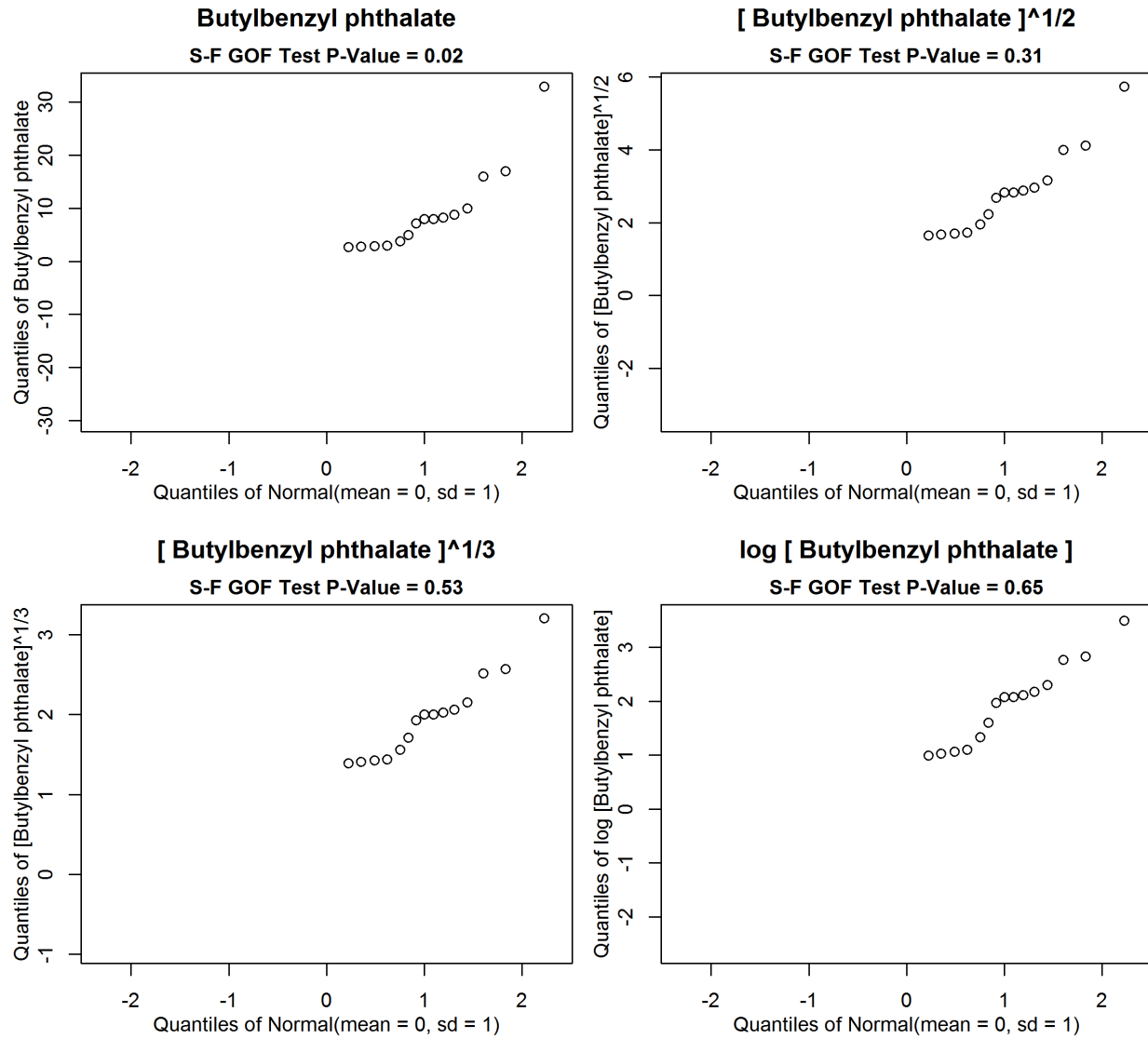
Figure 2.6: Censored Normal Q-Q Plots for Cadmium

Censored Normal Q-Q Plots for Cadmium (mg/kg)
Number of Censored Observations = 9 out of 67
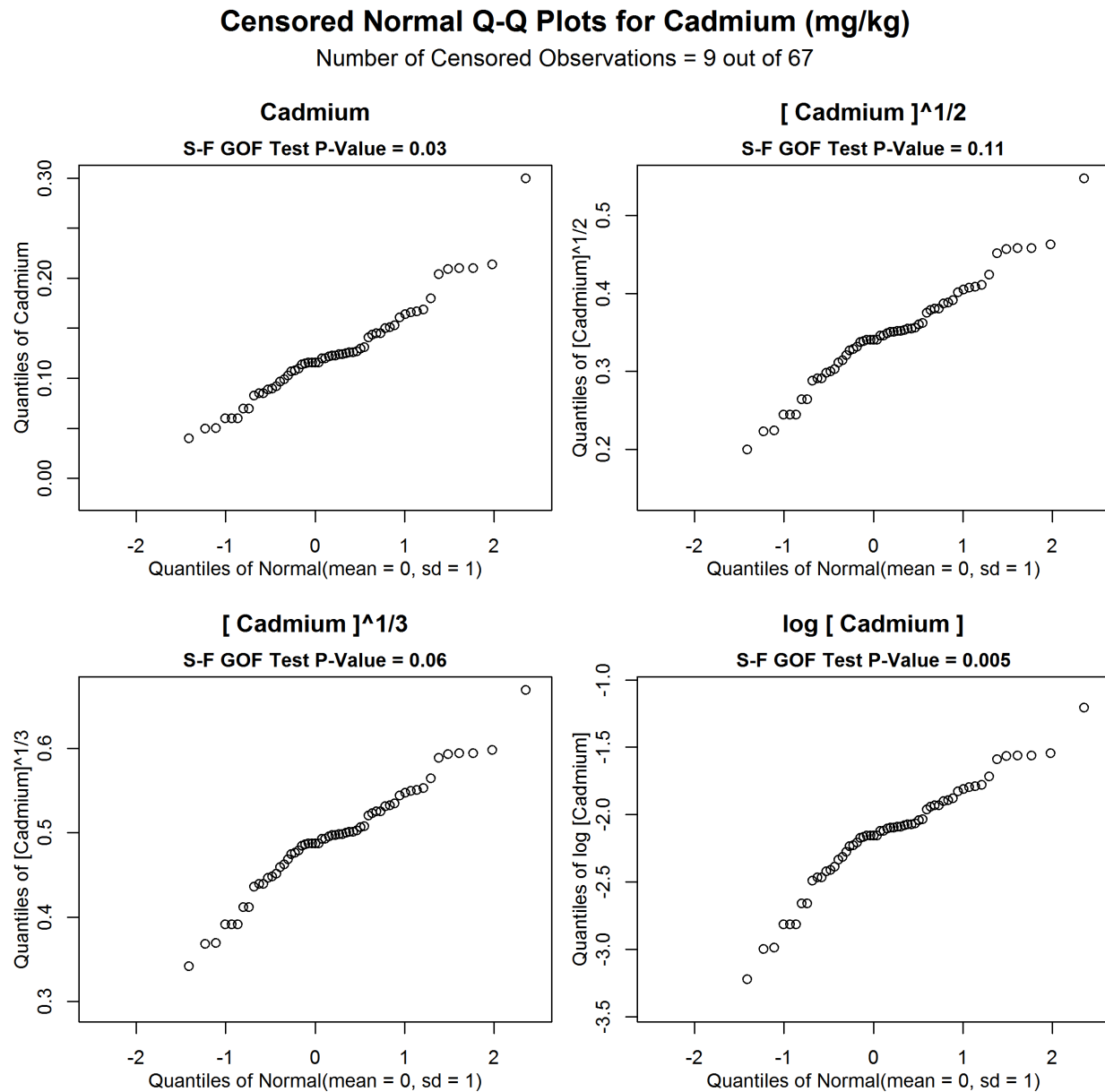
Figure 2.7: Normal Q-Q Plots for Chromium

Normal Q-Q Plots for Chromium (mg/kg)

Number of Observations = 65

Figure 2.8: Normal Q-Q Plots for Copper

Normal Q-Q Plots for Copper (mg/kg)
Number of Observations = 67

Figure 2.9: Censored Normal Q-Q Plots for CPAH

Censored Normal Q-Q Plots for CPAH (ug/kg)

Number of Censored Observations = 8 out of 52

Figure 2.10: Censored Normal Q-Q Plots for Dieldrin

Censored Normal Q-Q Plots for Dieldrin (ug/kg)

Number of Censored Observations = 41 out of 48
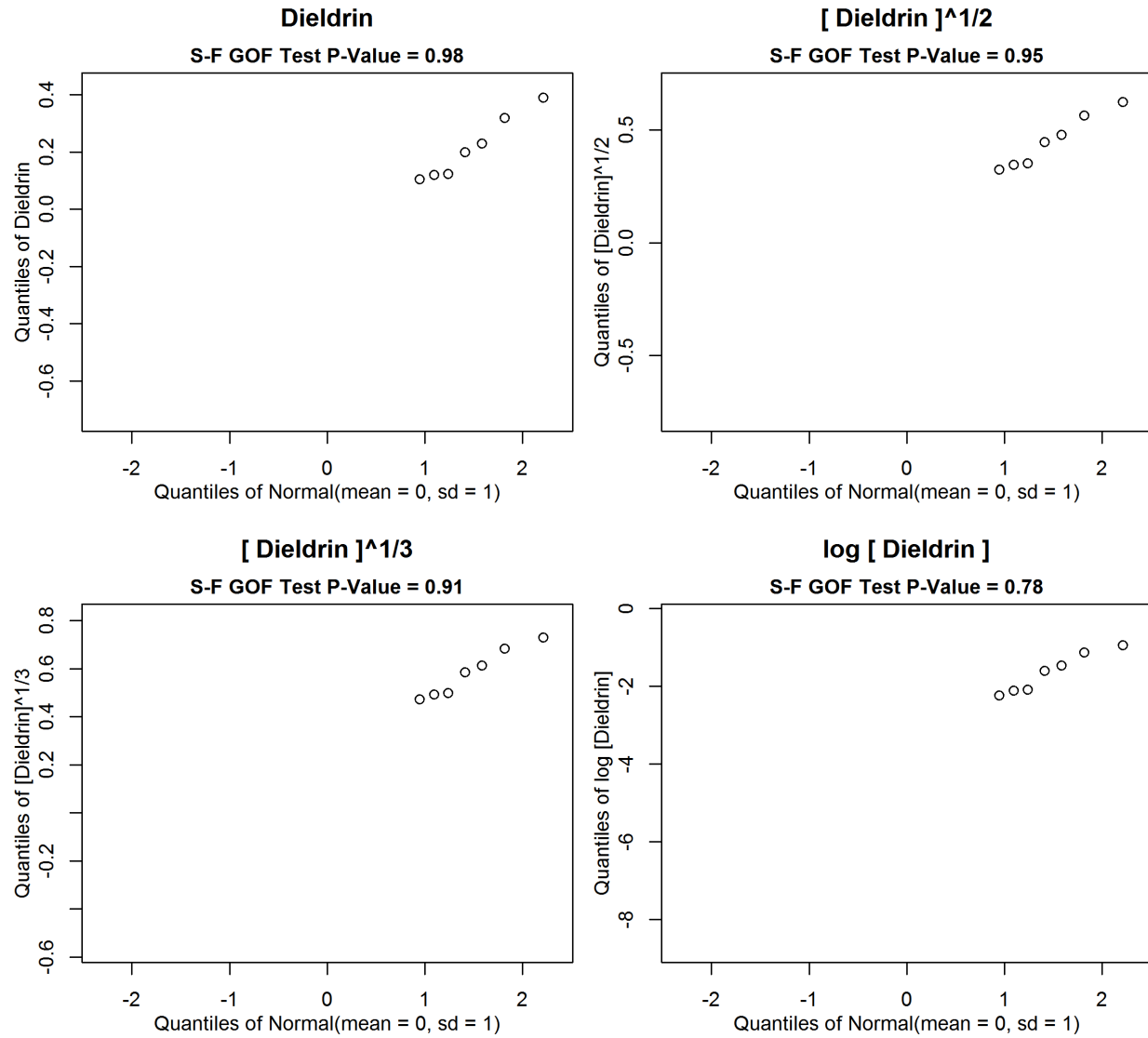
Figure 2.11: Normal Q-Q Plots for Diesel Range Hydrocarbons


Normal Q-Q Plots for Diesel Range Hydrocarbons (mg/kg)

Number of Observations = 28
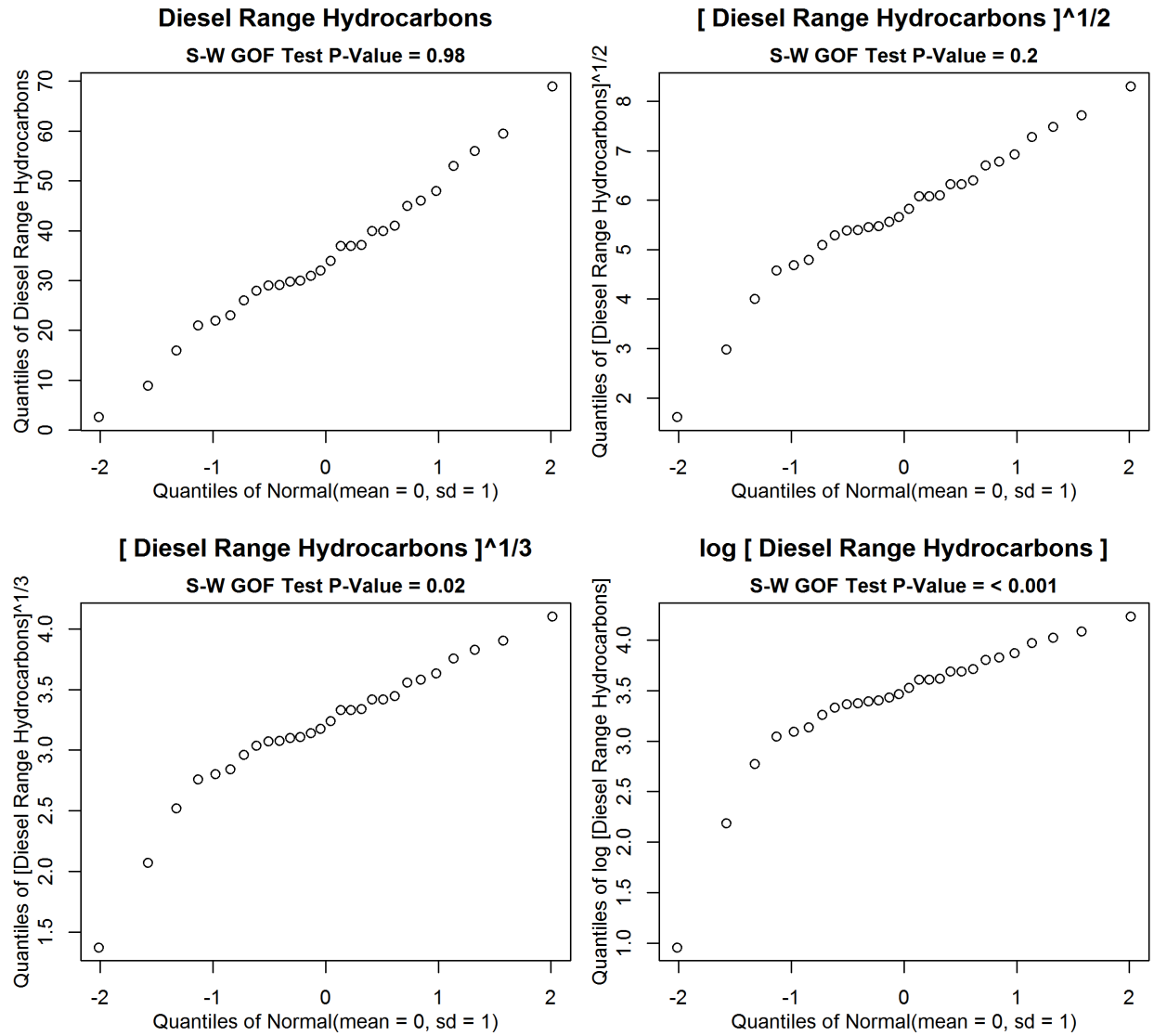
Figure 2.12: Normal Q-Q Plots for Dioxin TEQ – Mammals 2006

Normal Q-Q Plots for Dioxin TEQ - Mammals 2006 (pg/g)
Number of Observations = 52
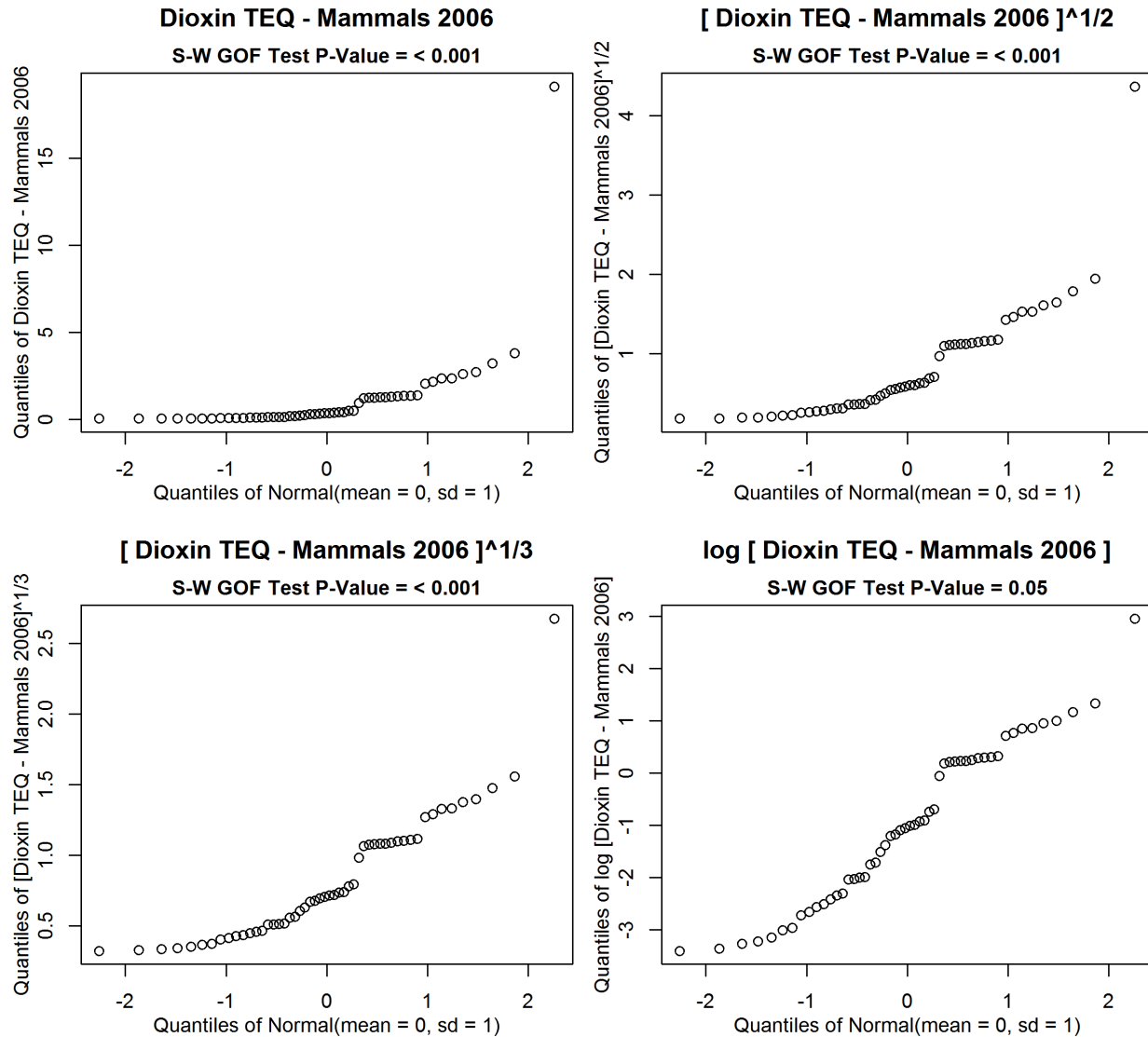
Figure 2.13: Censored Normal Q-Q Plots for gamma-Hexachlorocyclohexane

Censored Normal Q-Q Plots for gamma-Hexachlorocyclohexane (ug/kg)
Number of Censored Observations = 47 out of 48

Figure 2.14: Censored Normal Q-Q Plots for Hexachlorobenzene

Censored Normal Q-Q Plots for Hexachlorobenzene (ug/kg)
Number of Censored Observations = 24 out of 48
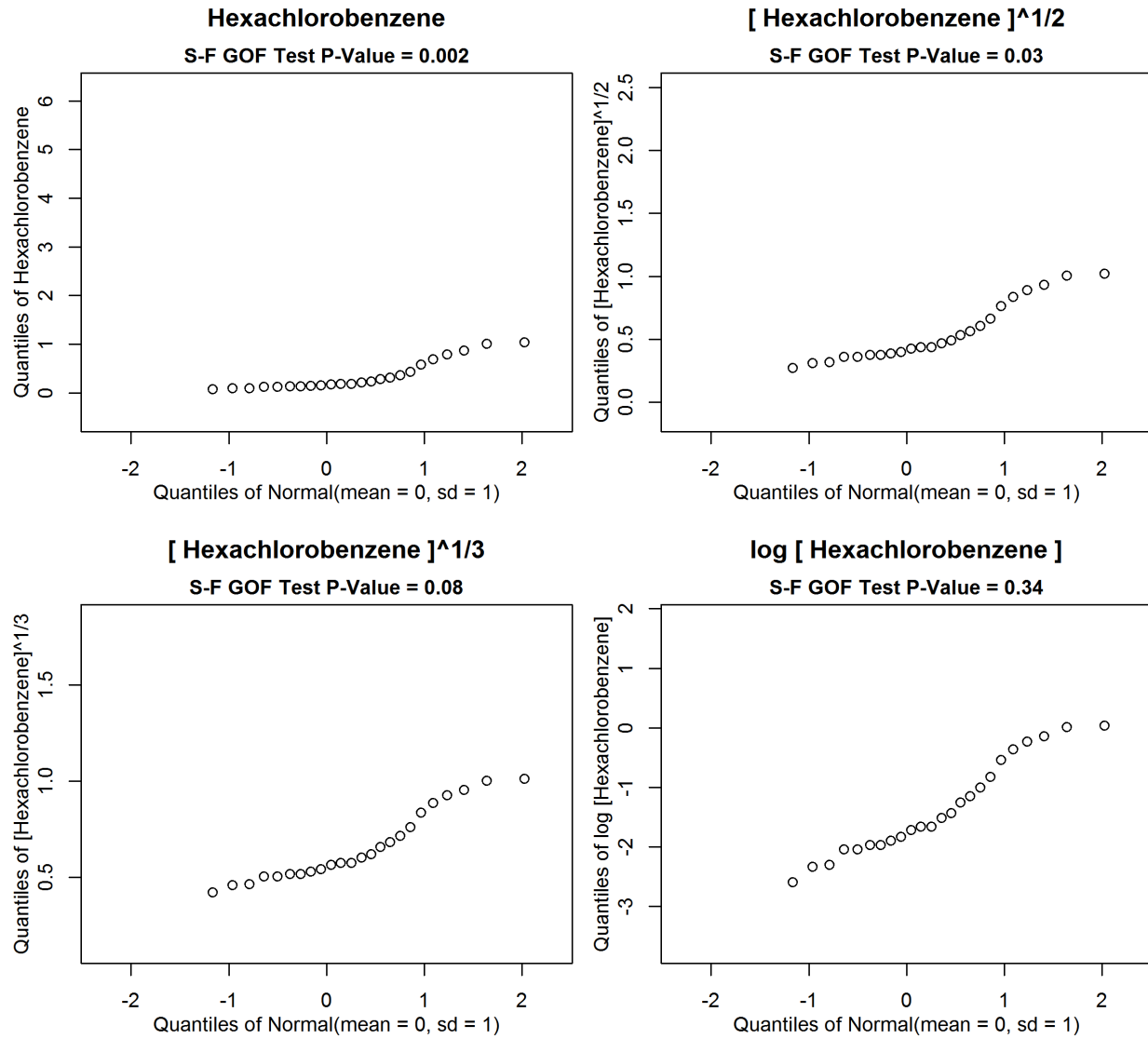
Figure 2.15: Normal Q-Q Plots for Lead

Normal Q-Q Plots for Lead (mg/kg)
Number of Observations = 67
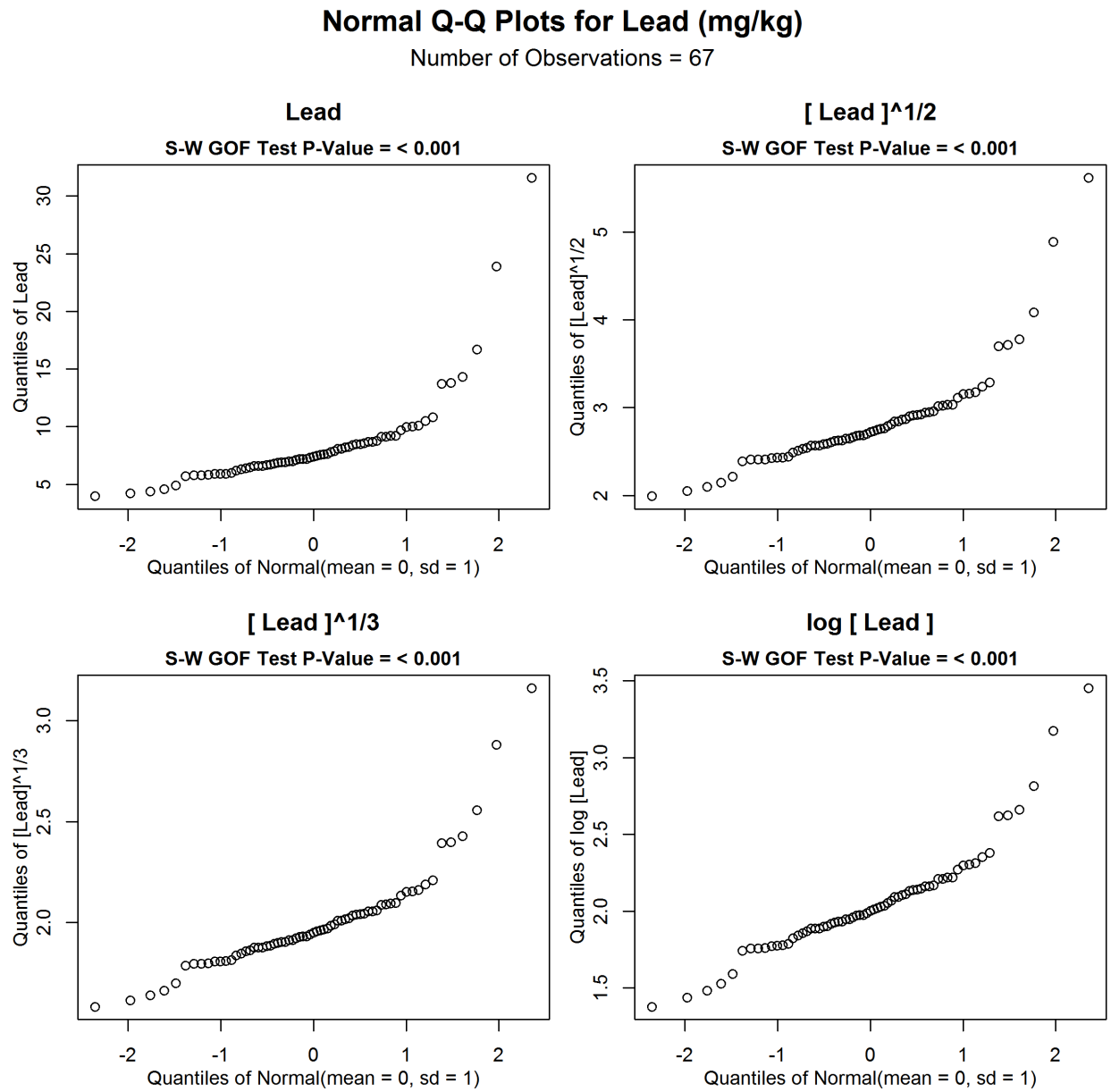
Figure 2.16: Censored Normal Q-Q Plots for Mercury

Censored Normal Q-Q Plots for Mercury (mg/kg)

Number of Censored Observations = 9 out of 61
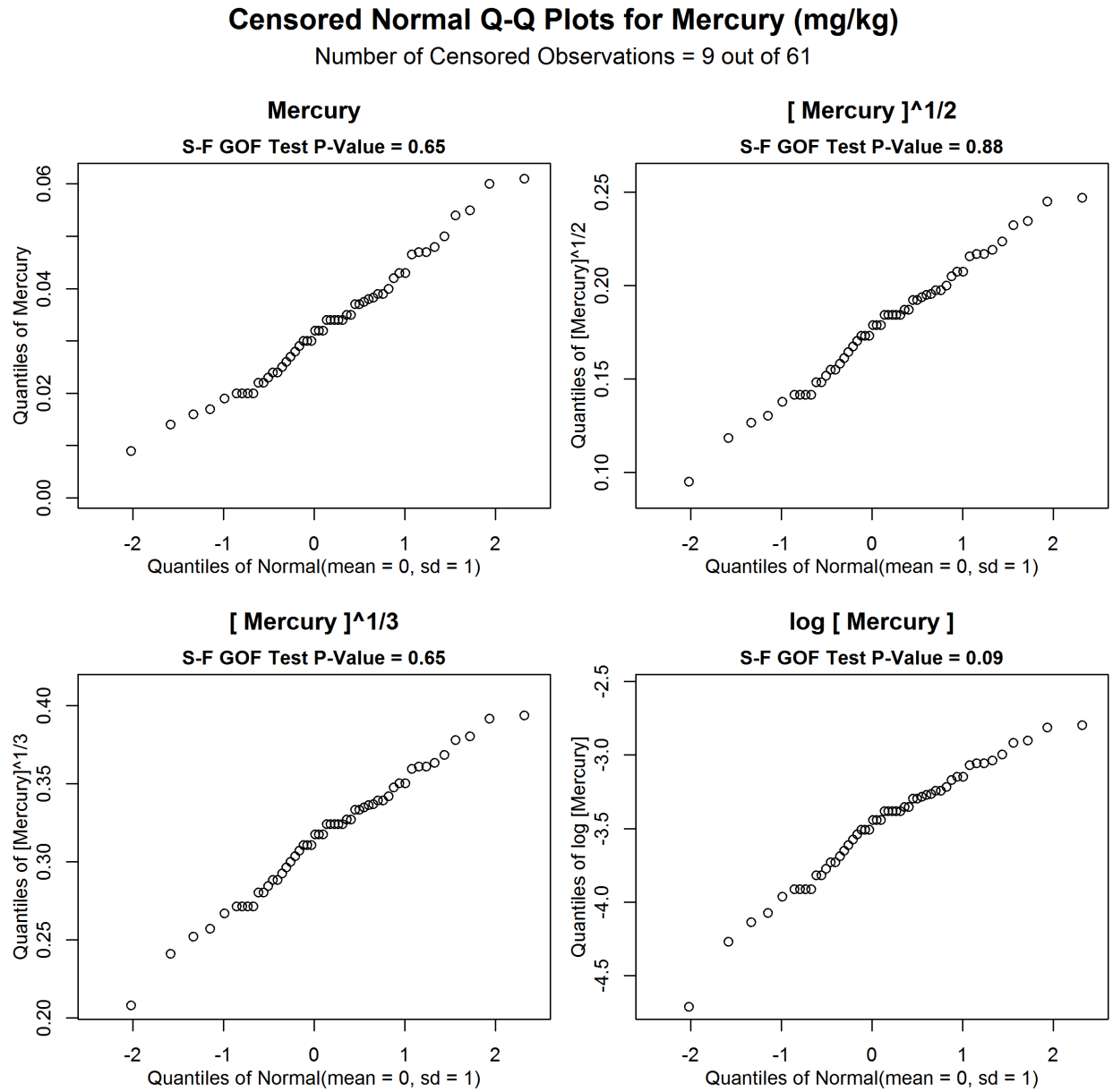
Figure 2.17: Censored Normal Q-Q Plots for Naphthalene

Censored Normal Q-Q Plots for Naphthalene (ug/kg)

Number of Censored Observations = 33 out of 52
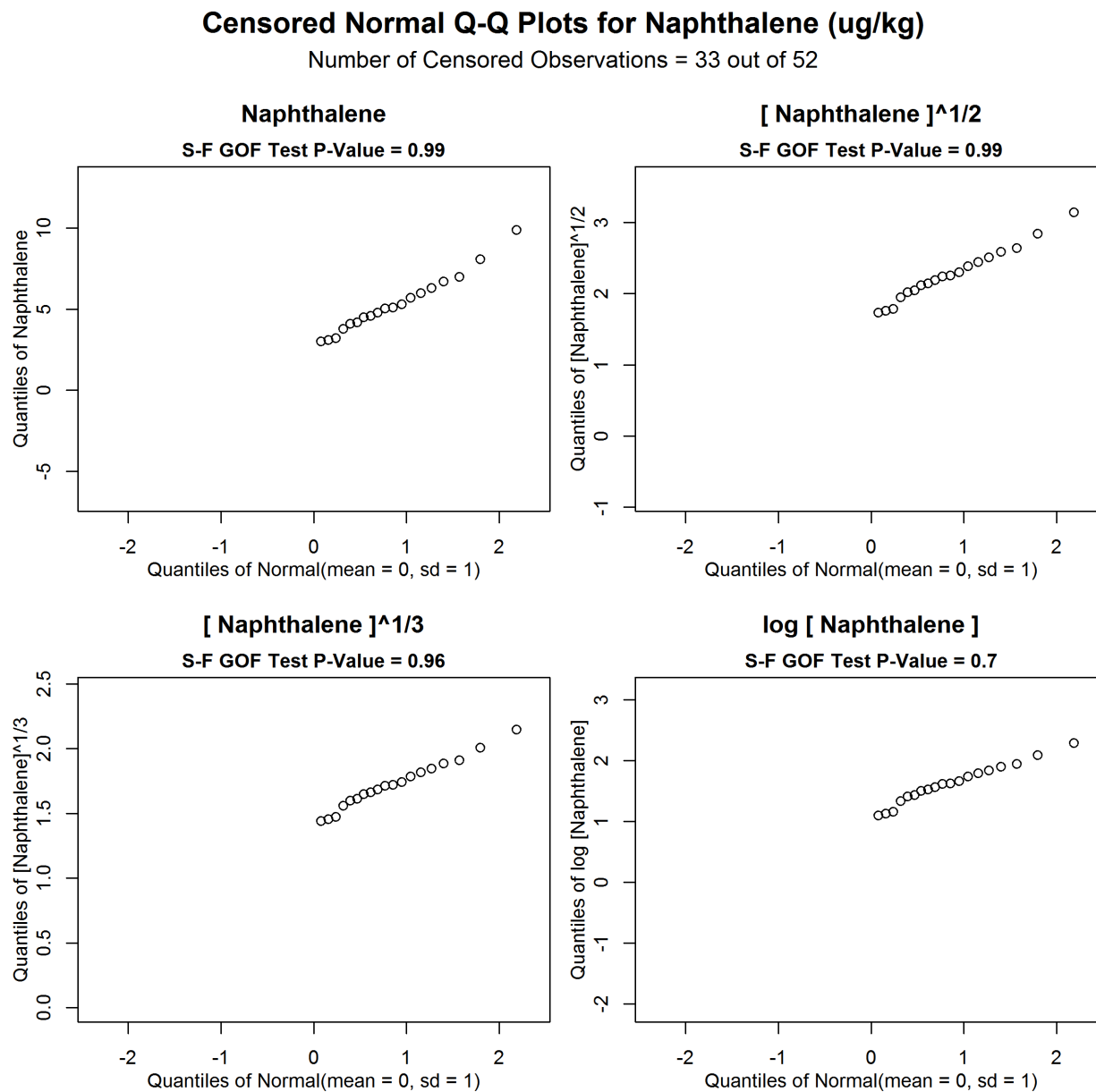
Figure 2.18: Normal Q-Q Plots for Nickel

Normal Q-Q Plots for Nickel (mg/kg)

Number of Observations = 67
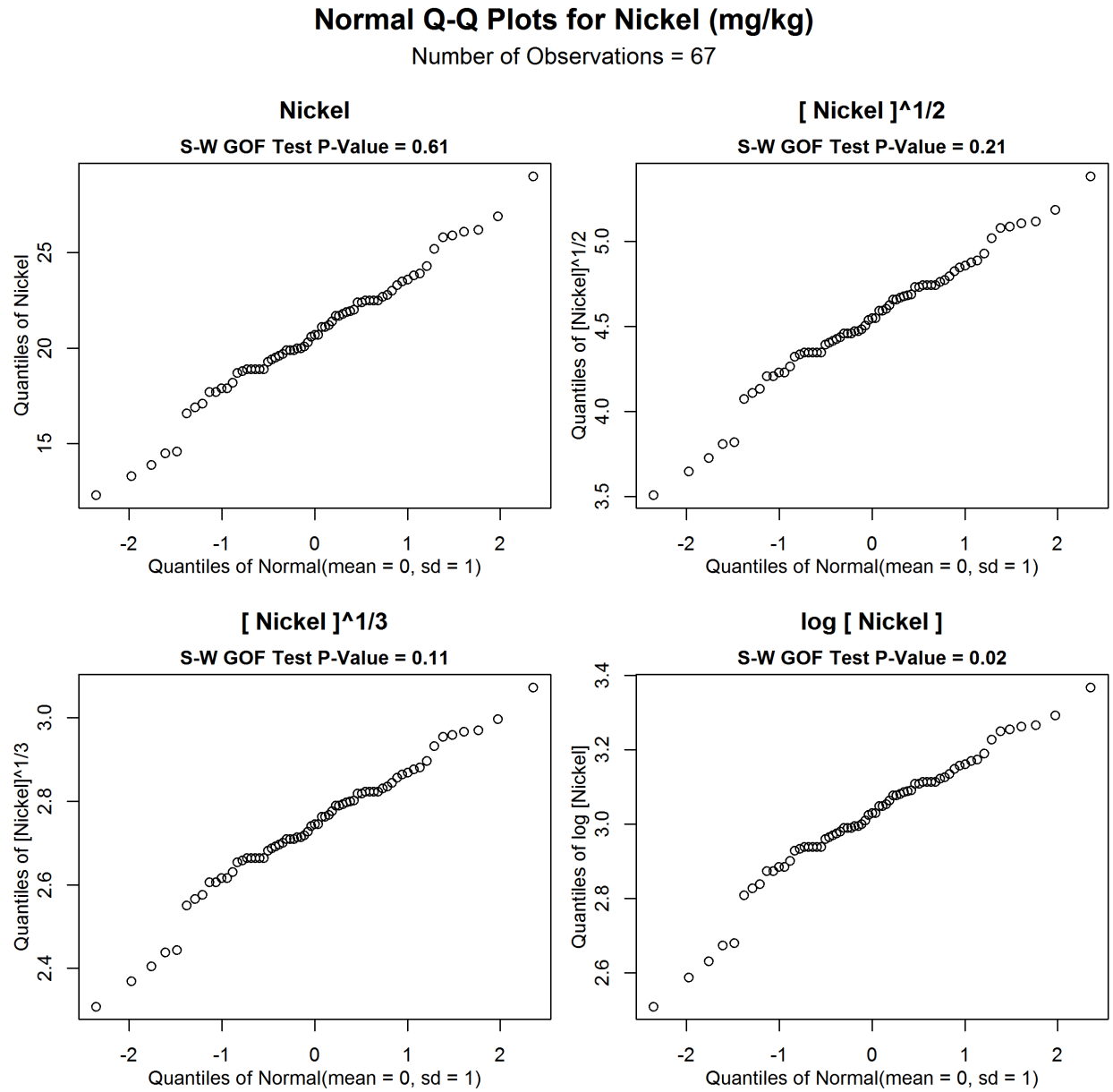
Figure 2.19: Censored Normal Q-Q Plots for PCB TEQ – Mammals 2006

Censored Normal Q-Q Plots for PCB TEQ - Mammals 2006 (pg/g)
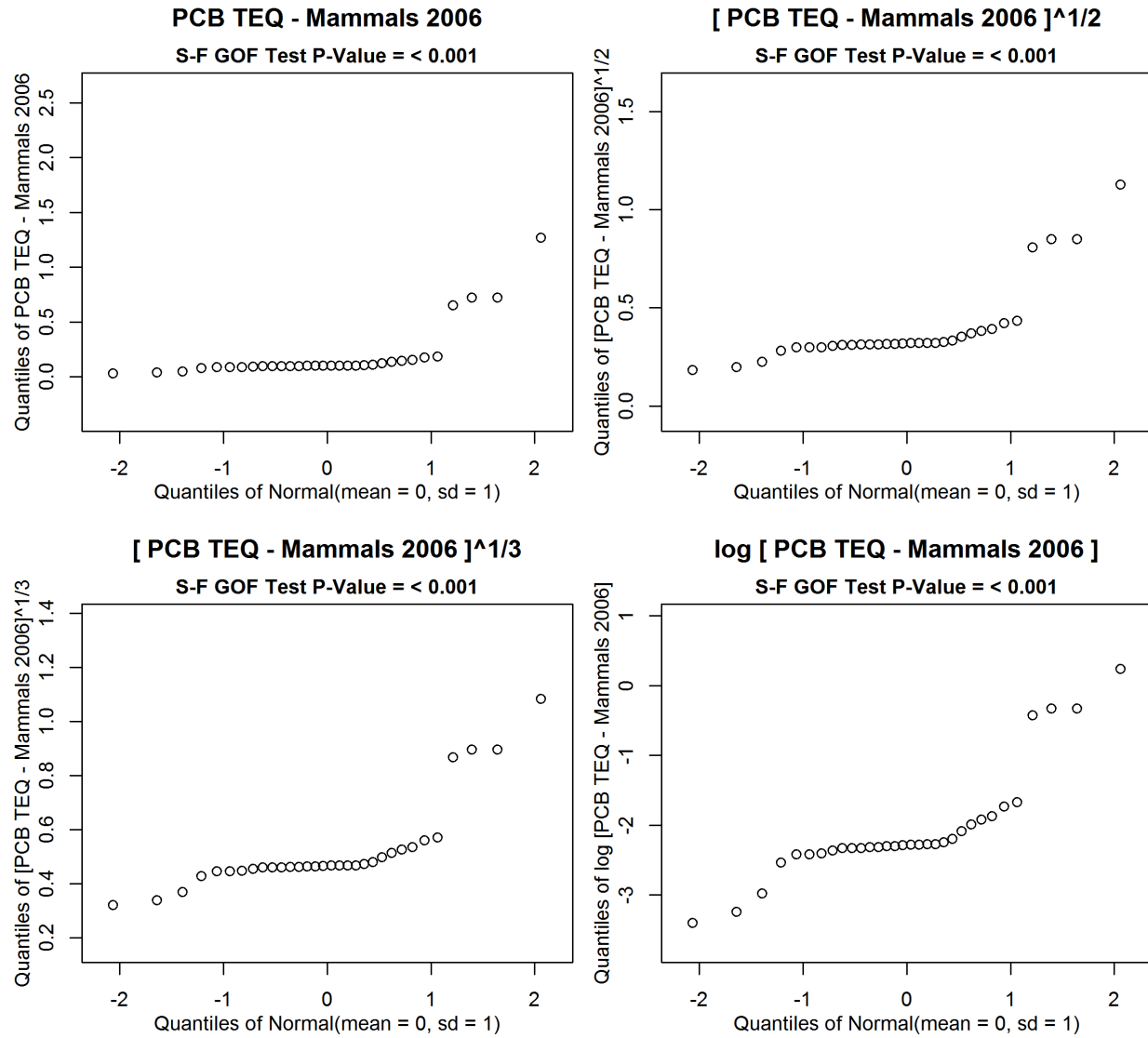
Number of Censored Observations = 1 out of 33

Censored Normal Q-Q Plots for Pentachlorophenol (ug/kg)
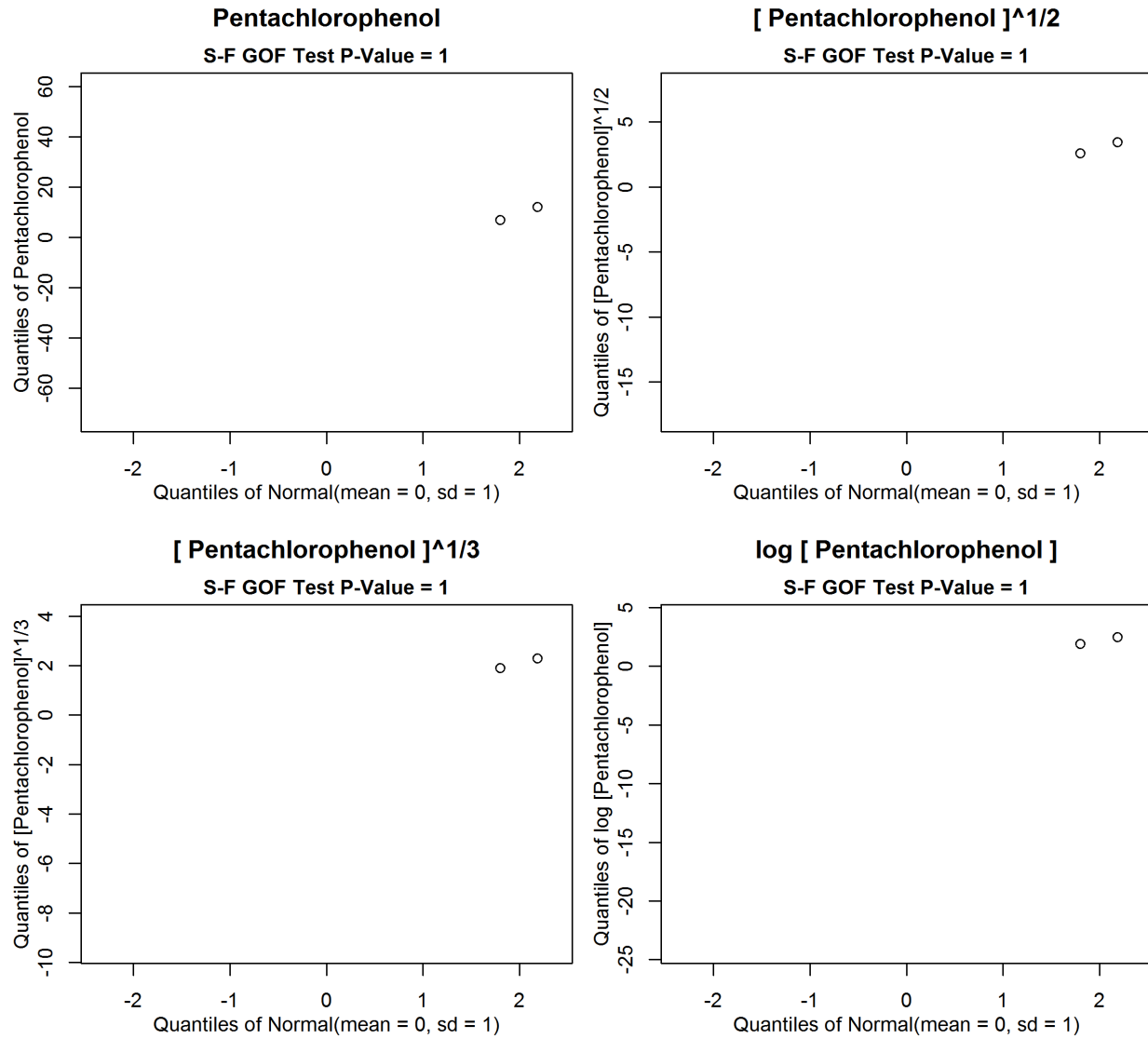
Number of Censored Observations = 50 out of 52

Figure 2.21: Censored Normal Q-Q Plots for Phenanthrene

Censored Normal Q-Q Plots for Phenanthrene (ug/kg)
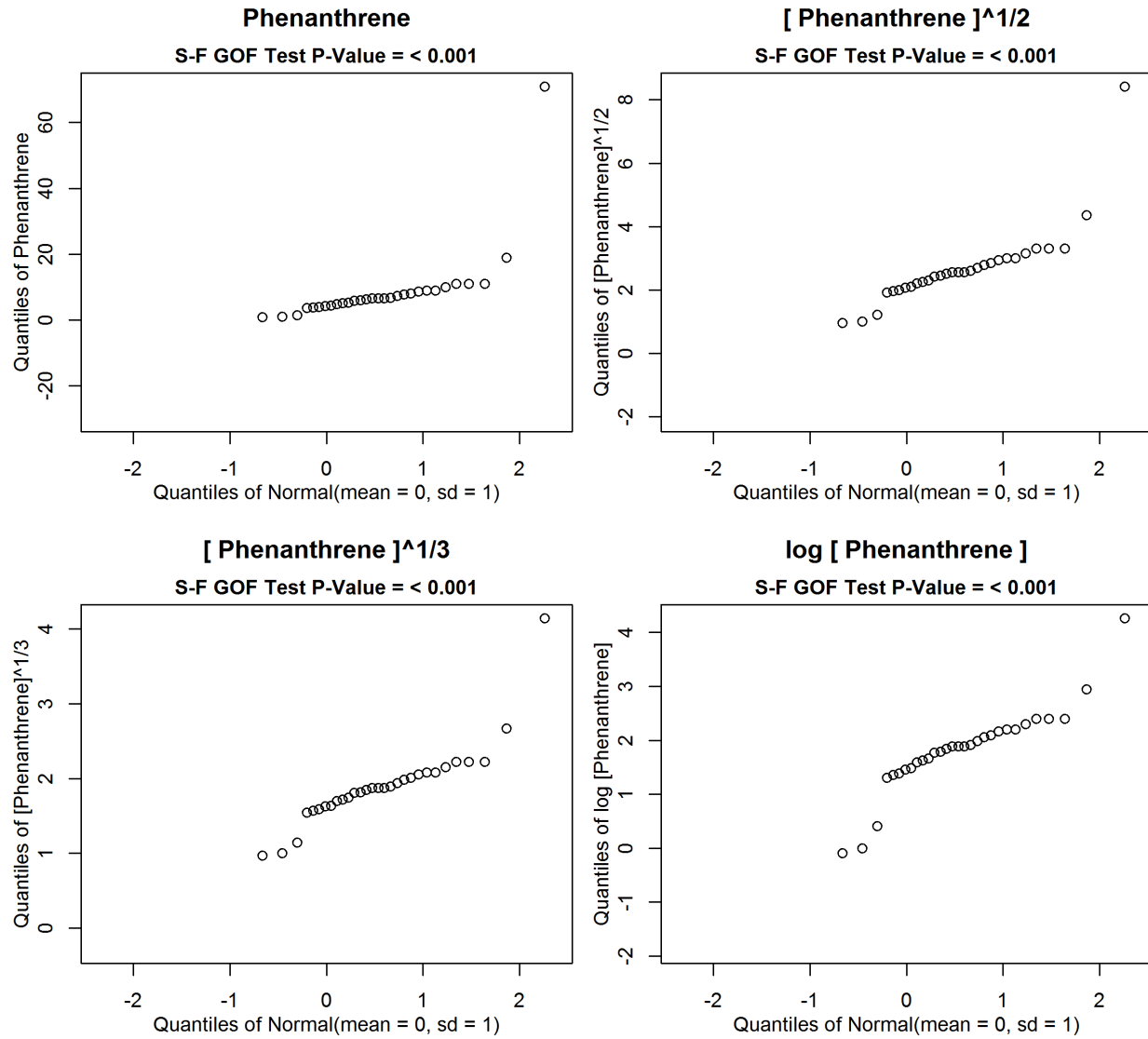
Number of Censored Observations = 22 out of 52

Figure 2.22: Normal Q-Q Plots for Residual Range Hydrocarbons

Normal Q-Q Plots for Residual Range Hydrocarbons (mg/kg)
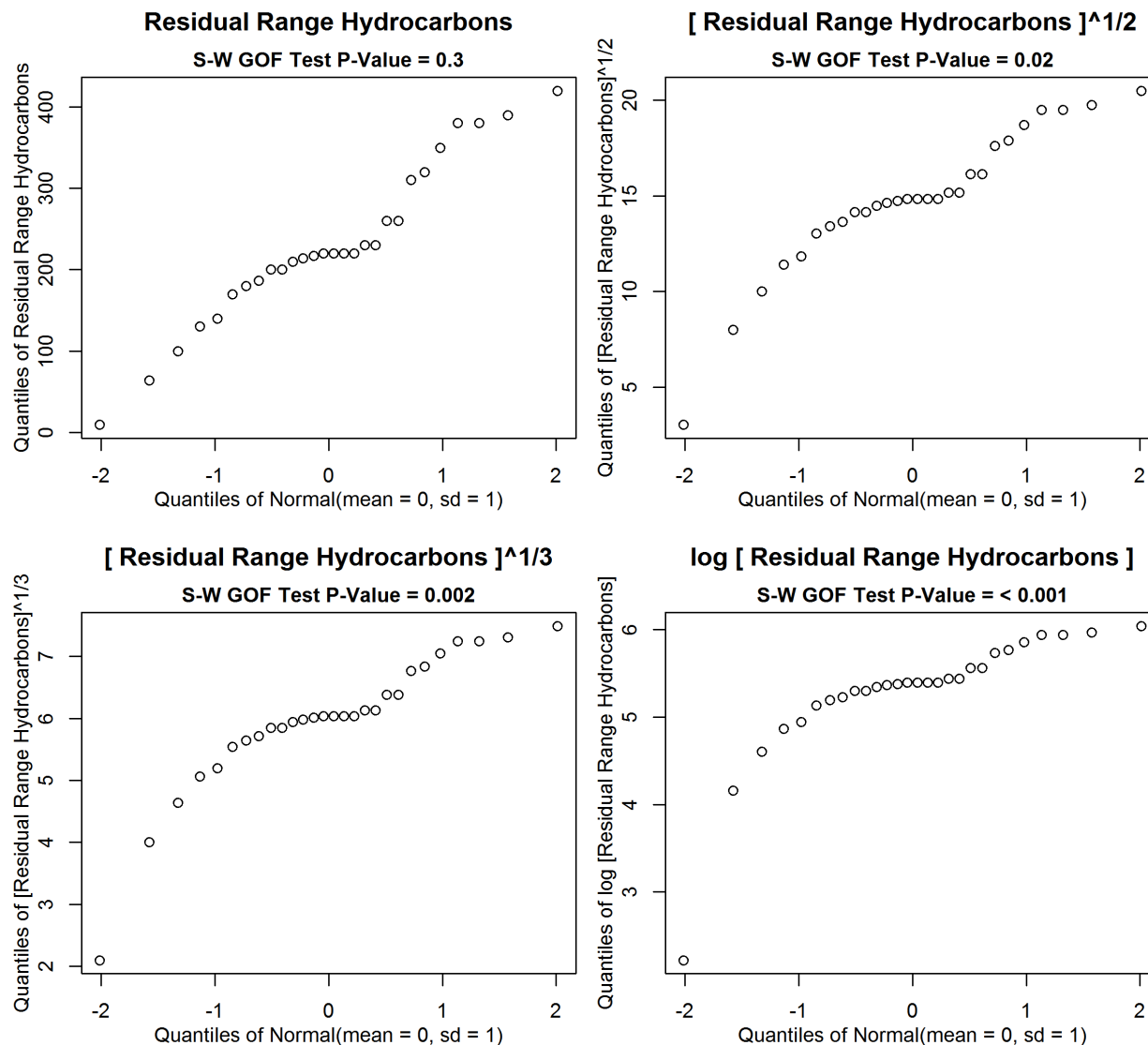
Number of Observations = 28

Figure 2.23: Censored Normal Q-Q Plots for Total Chlordane

Censored Normal Q-Q Plots for Total Chlordane (ug/kg)
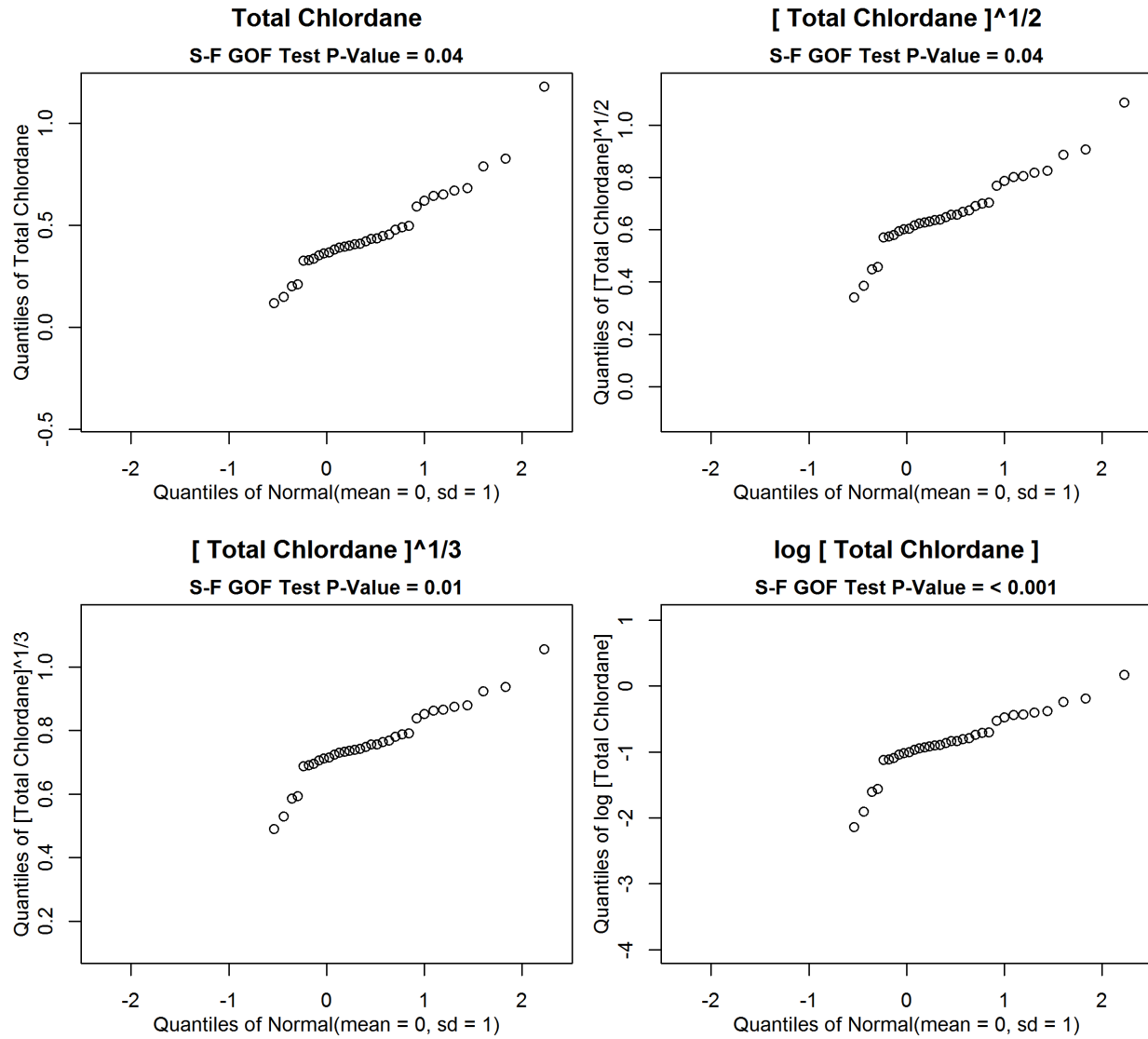
Number of Censored Observations = 15 out of 48

Total Chlordane

S-F GOF Test P-Value = 0.04

[Total Chlordane]^1/2

S-F GOF Test P-Value = 0.04

[Total Chlordane]^1/3

S-F GOF Test P-Value = 0.01

log [Total Chlordane]

S-F GOF Test P-Value = < 0.001

Censored Normal Q-Q Plots for Total DDTs (ug/kg)

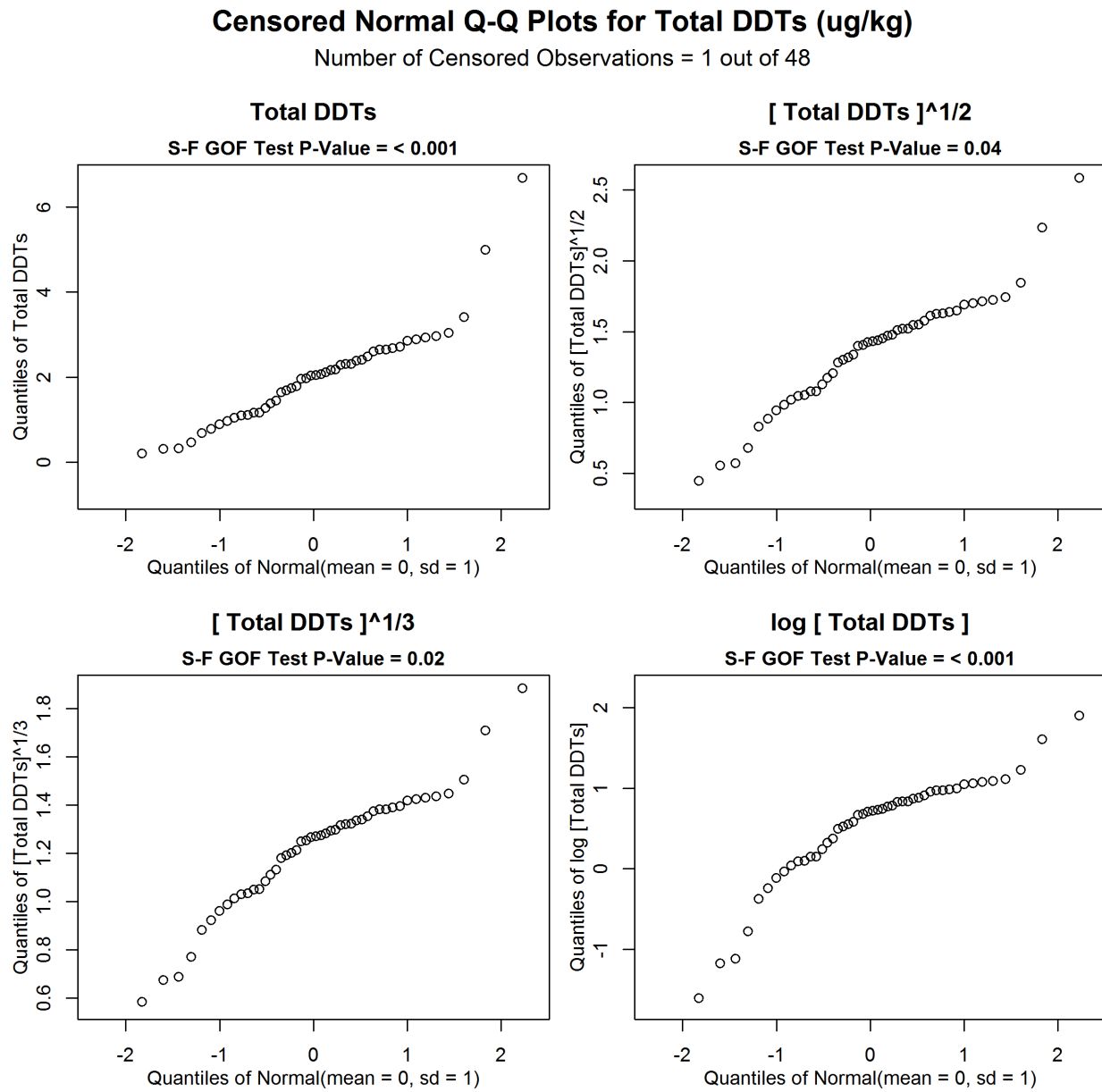Number of Censored Observations = 1 out of 48

Figure 2.25: Censored Normal Q-Q Plots for Total HPAHs

Censored Normal Q-Q Plots for Total HPAHs (ug/kg)

Number of Censored Observations = 8 out of 52

Total HPAHs — S-F GOF Test P-Value = < 0.001

[Total HPAHs]^1/2 — S-F GOF Test P-Value = 0.005

[Total HPAHs]^1/3 — S-F GOF Test P-Value = 0.02

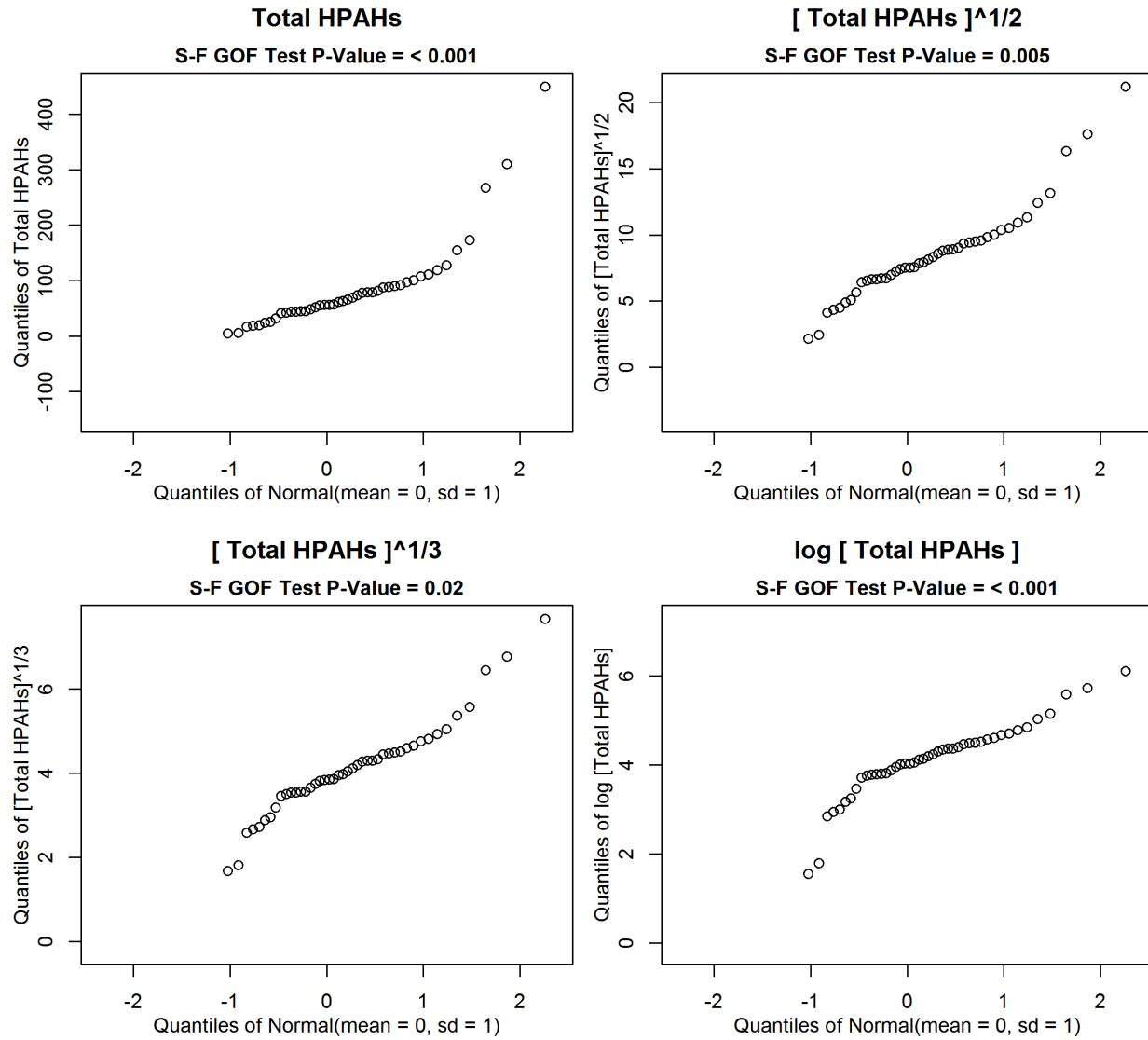log [Total HPAHs] — S-F GOF Test P-Value = < 0.001

Figure 2.26: Censored Normal Q-Q Plots for Total LPAHs

Censored Normal Q-Q Plots for Total LPAHs (ug/kg)

Number of Censored Observations = 12 out of 52

Figure 2.27: Censored Normal Q-Q Plots for Total PAHs

Censored Normal Q-Q Plots for Total PAHs (ug/kg)

Number of Censored Observations = 7 out of 52
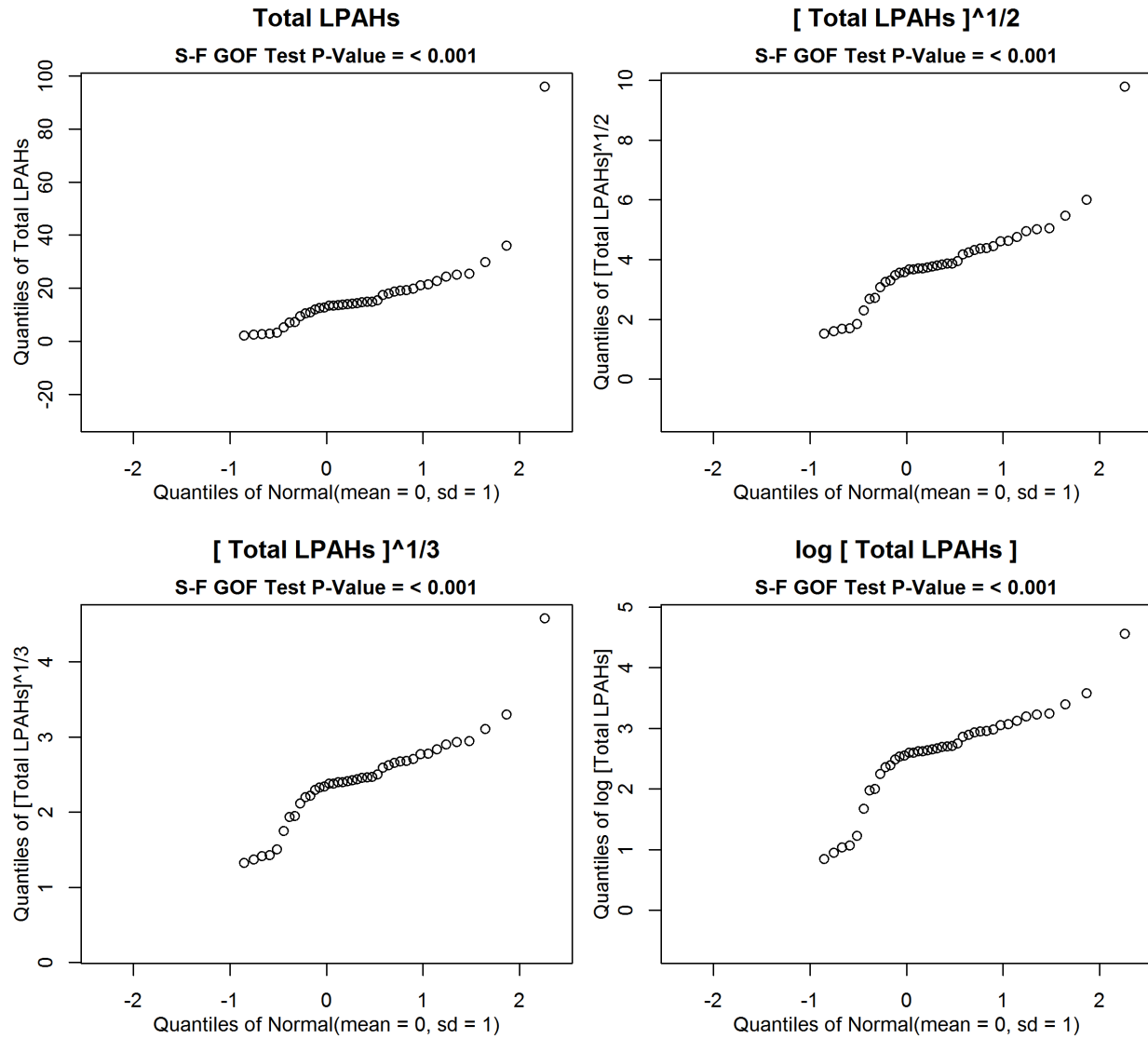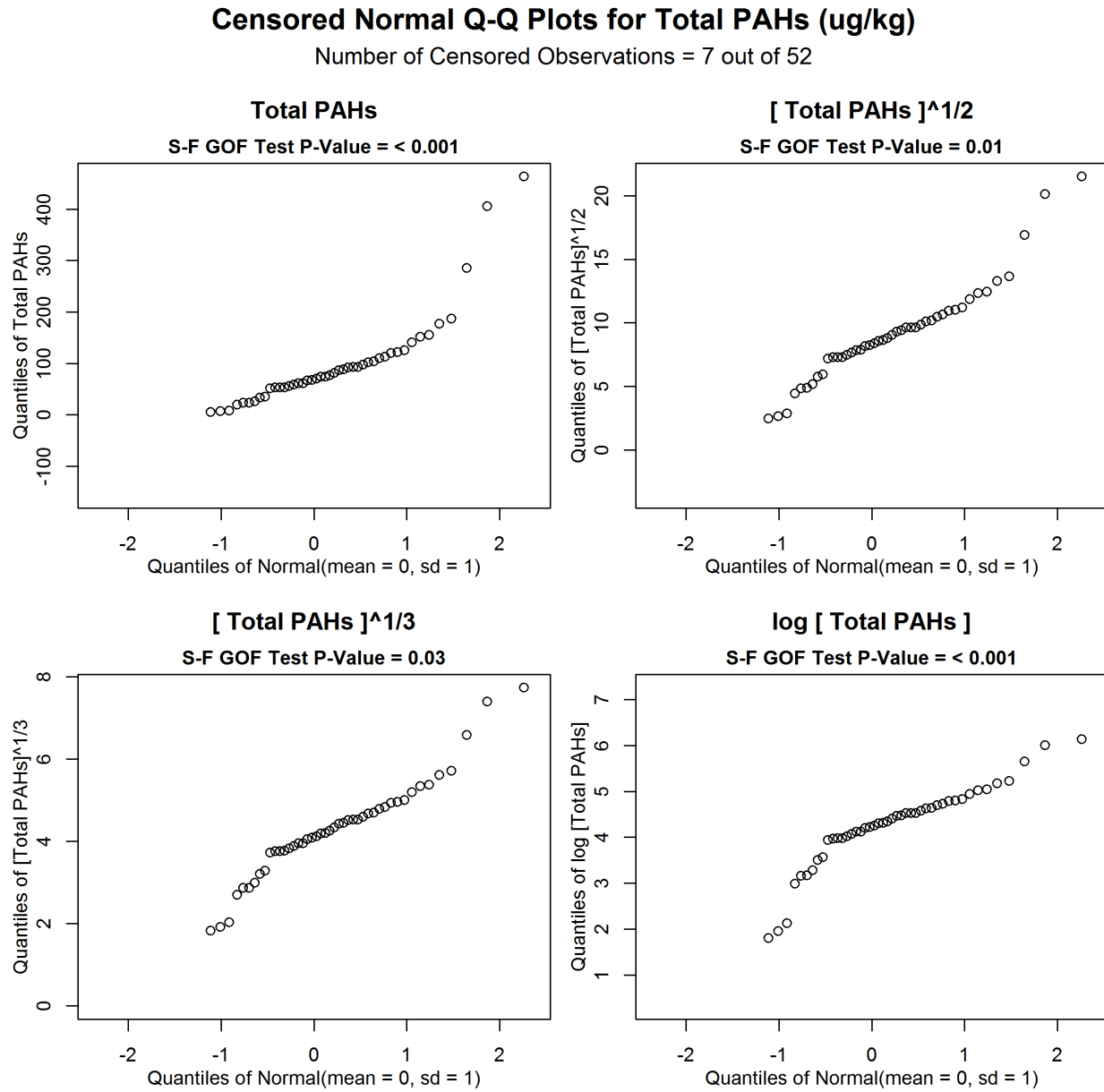
Figure 2.28: Normal Q-Q Plots for Total PCB Congeners

Normal Q-Q Plots for Total PCB Congeners (pg/g)
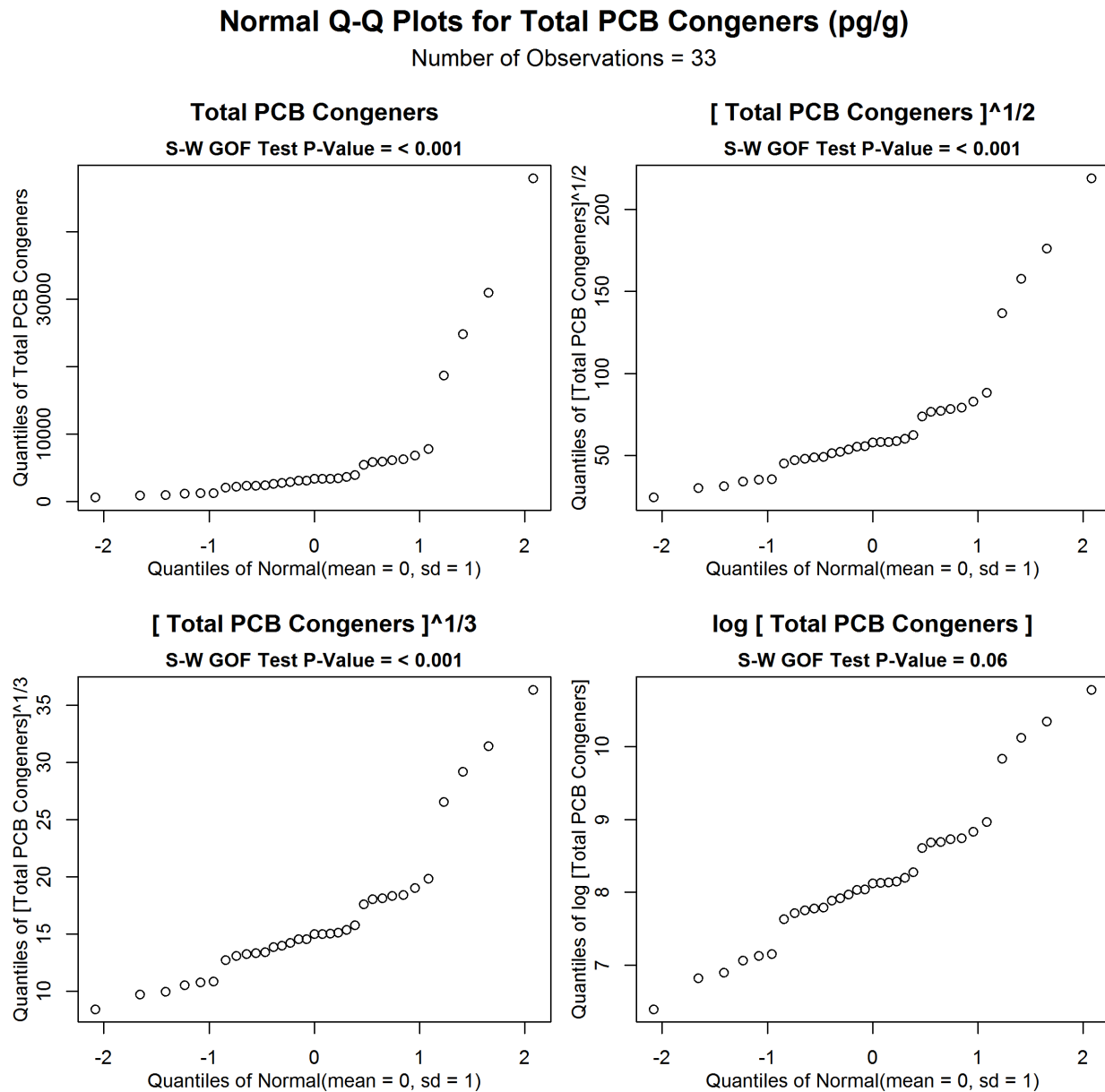
Number of Observations = 33

Figure 2.29: Censored Normal Q-Q Plots for Total PCBs Aroclors

Censored Normal Q-Q Plots for Total PCBs Aroclors (ug/kg)
Number of Censored Observations = 25 out of 48

Total PCBs Aroclors
S-F GOF Test P-Value = 0.009

[Total PCBs Aroclors]^1/2
S-F GOF Test P-Value = 0.1

[Total PCBs Aroclors]^1/3
S-F GOF Test P-Value = 0.19

log [Total PCBs Aroclors]
S-F GOF Test P-Value = 0.49
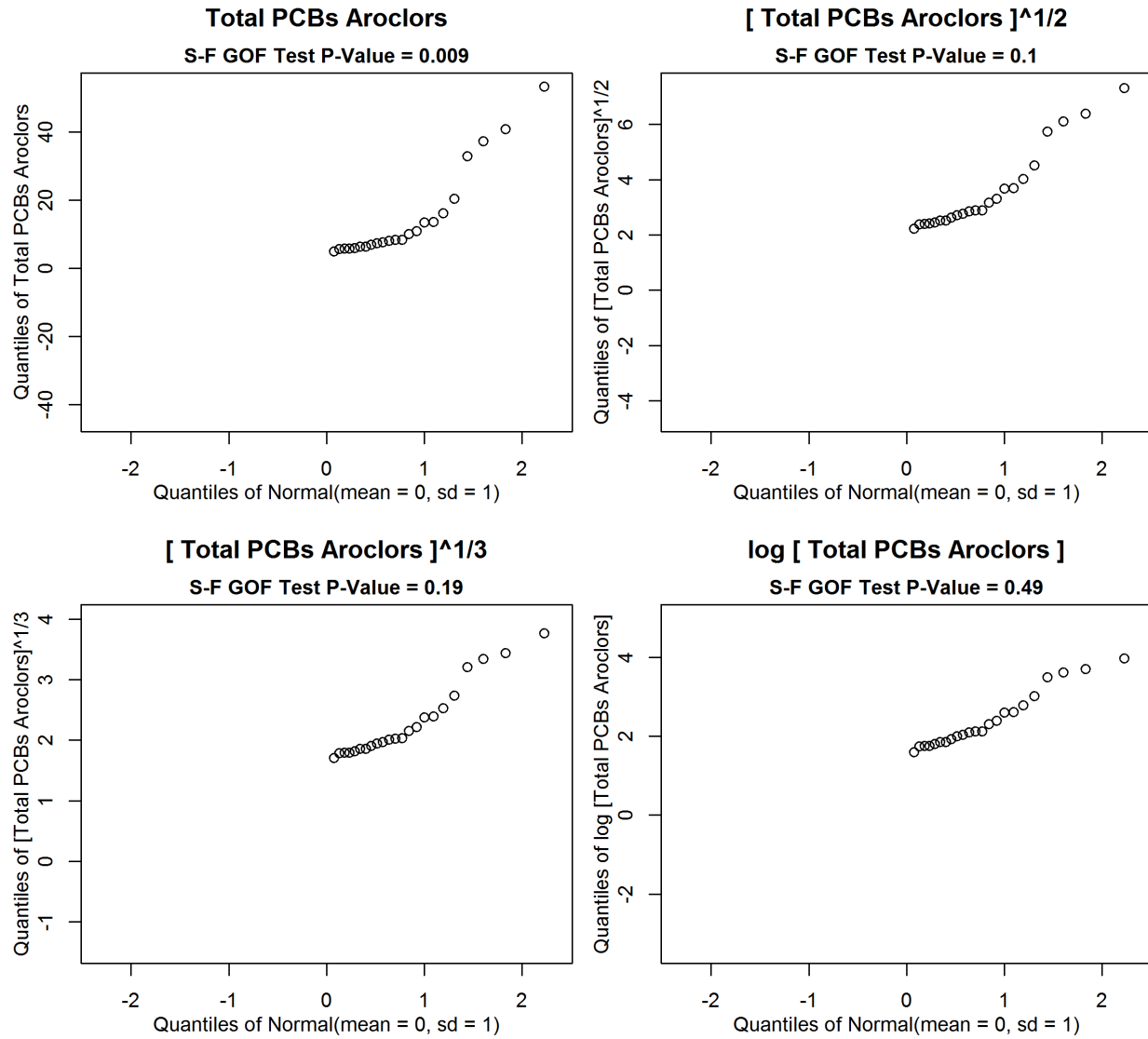
Figure 2.30: Normal Q-Q Plots for Total PCDD/F

Normal Q-Q Plots for Total PCDD/F (pg/g)

Number of Observations = 33
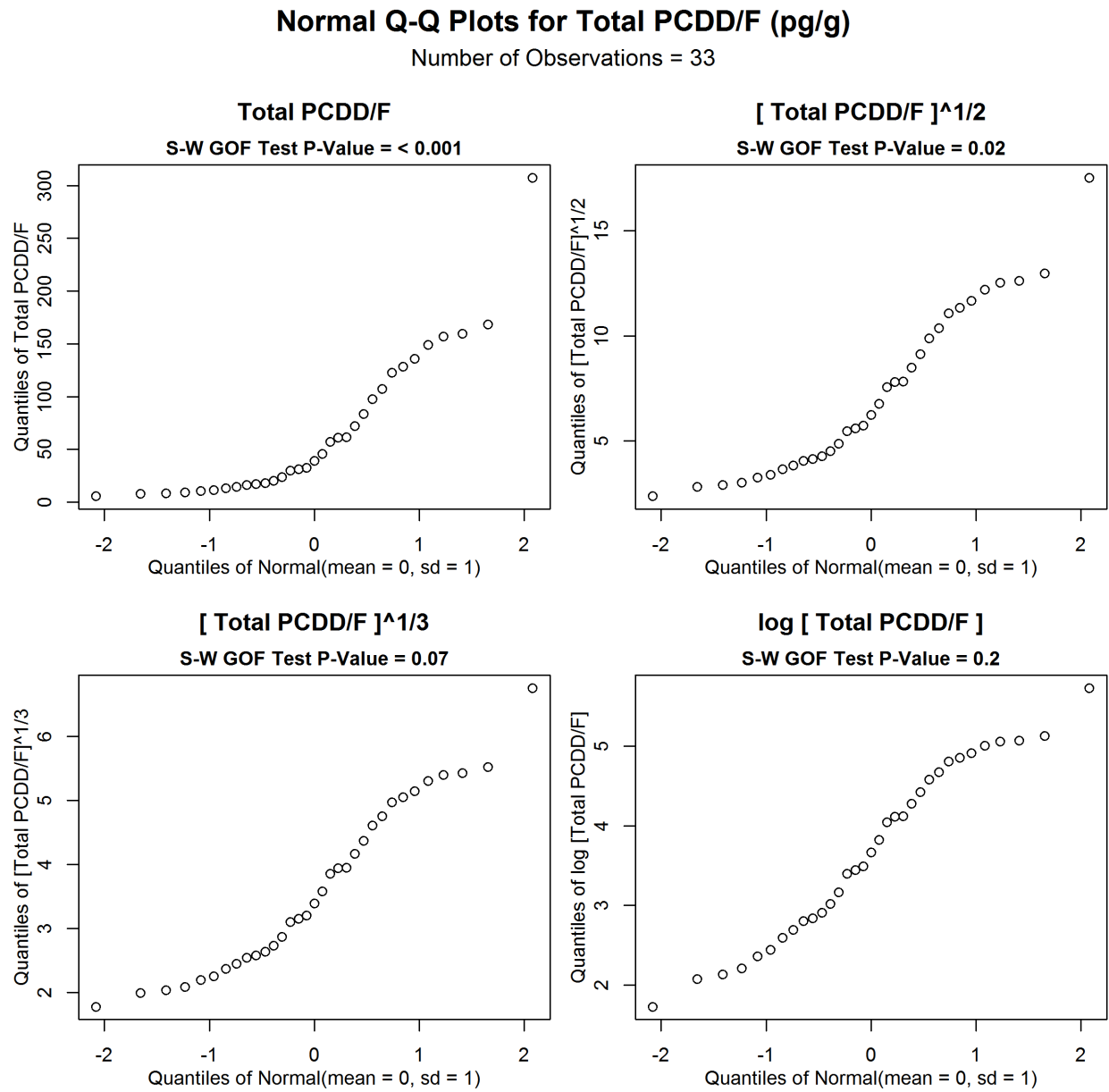
Figure 2.31: Normal Q-Q Plots for Total Petroleum Hydrocarbons

Normal Q-Q Plots for Total Petroleum Hydrocarbons (mg/kg)

Number of Observations = 28

Figure 2.32: Censored Normal Q-Q Plots for Tributyltin ion

Censored Normal Q-Q Plots for Tributyltin ion (ug/kg)

Number of Censored Observations = 1 out of 3
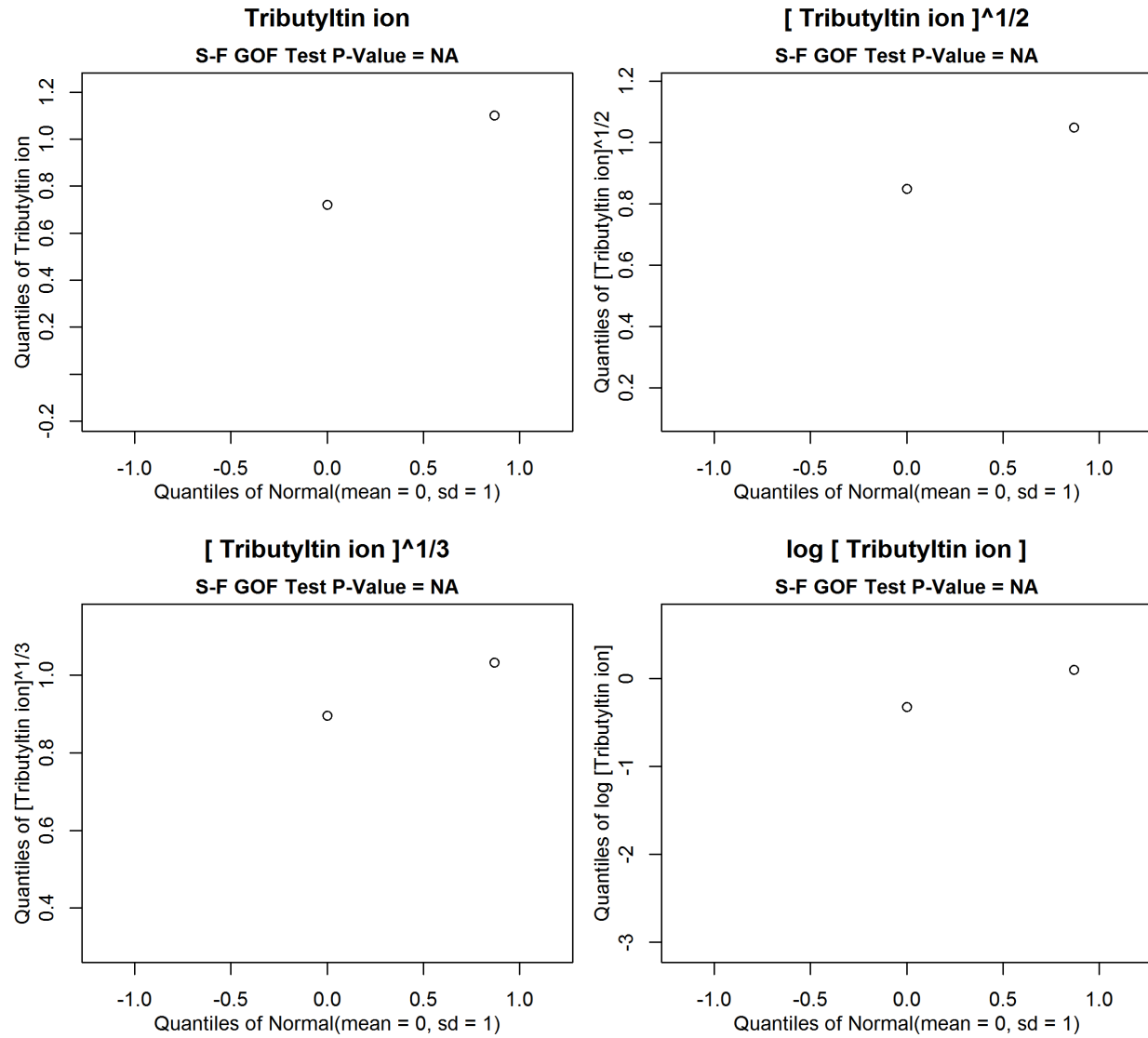
Figure 2.33: Normal Q-Q Plots for Zinc

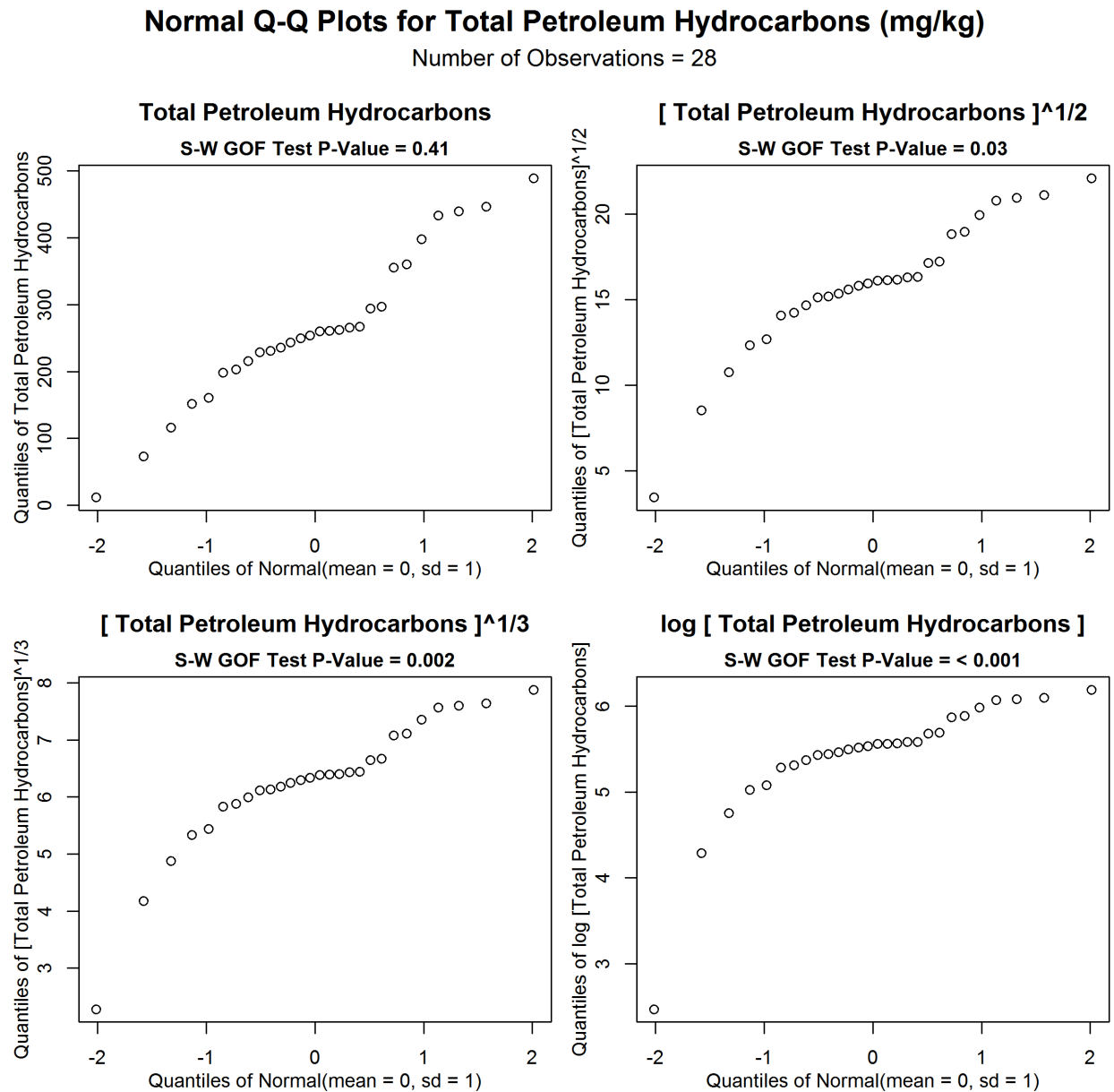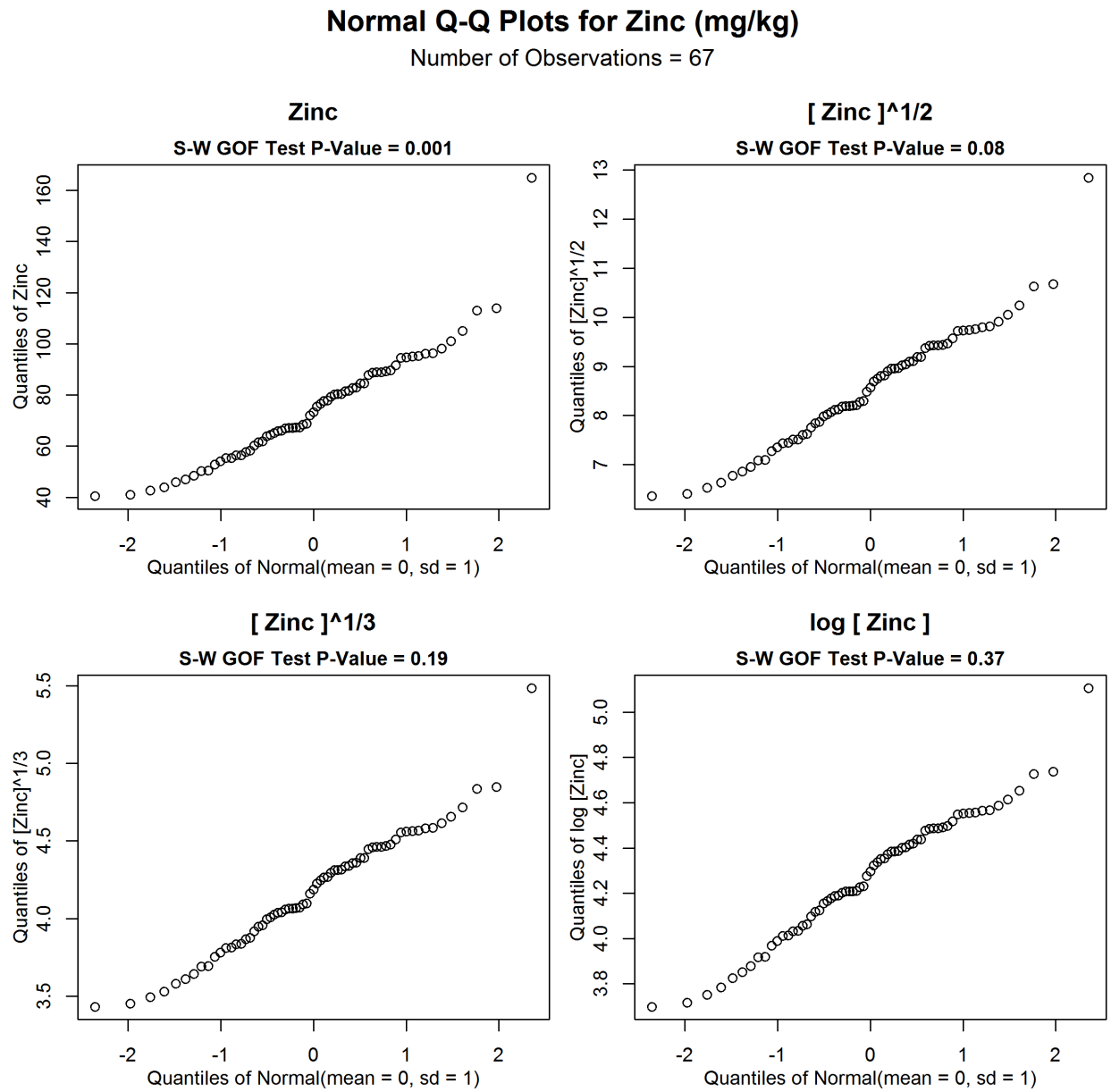Normal Q-Q Plots for Zinc (mg/kg)

Number of Observations = 67

Figure 3.1: Distribution of Number of Outliers According to Rosner's Test Based on Normal Distribution and Various Censoring Levels

Distribution of Number of Outliers According to Rosner's Test
Based on Singly Censored Data from a
Normal(Mean = 10, SD = 2) Distribution
k = 10; Assumed Type I Error = 5%; Number of Simulations = 1,000

Figure 3.2: Distribution of Number of Outliers According to Rosner's Test Based on Gamma Distribution and Various Censoring Levels

Distribution of Number of Outliers According to Rosner's Test
Based on Singly Censored Data from a
Gamma(Mean = 10, CV = 1) Distribution
k = 10; Assumed Type I Error = 5%; Number of Simulations = 1,000

15% Single Censoring
95% CI for Pr(At Least One Outlier) = [0.74, 0.8]

30% Single Censoring
95% CI for Pr(At Least One Outlier) = [0.75, 0.8]

45% Single Censoring
95% CI for Pr(At Least One Outlier) = [0.78, 0.83]

60% Single Censoring
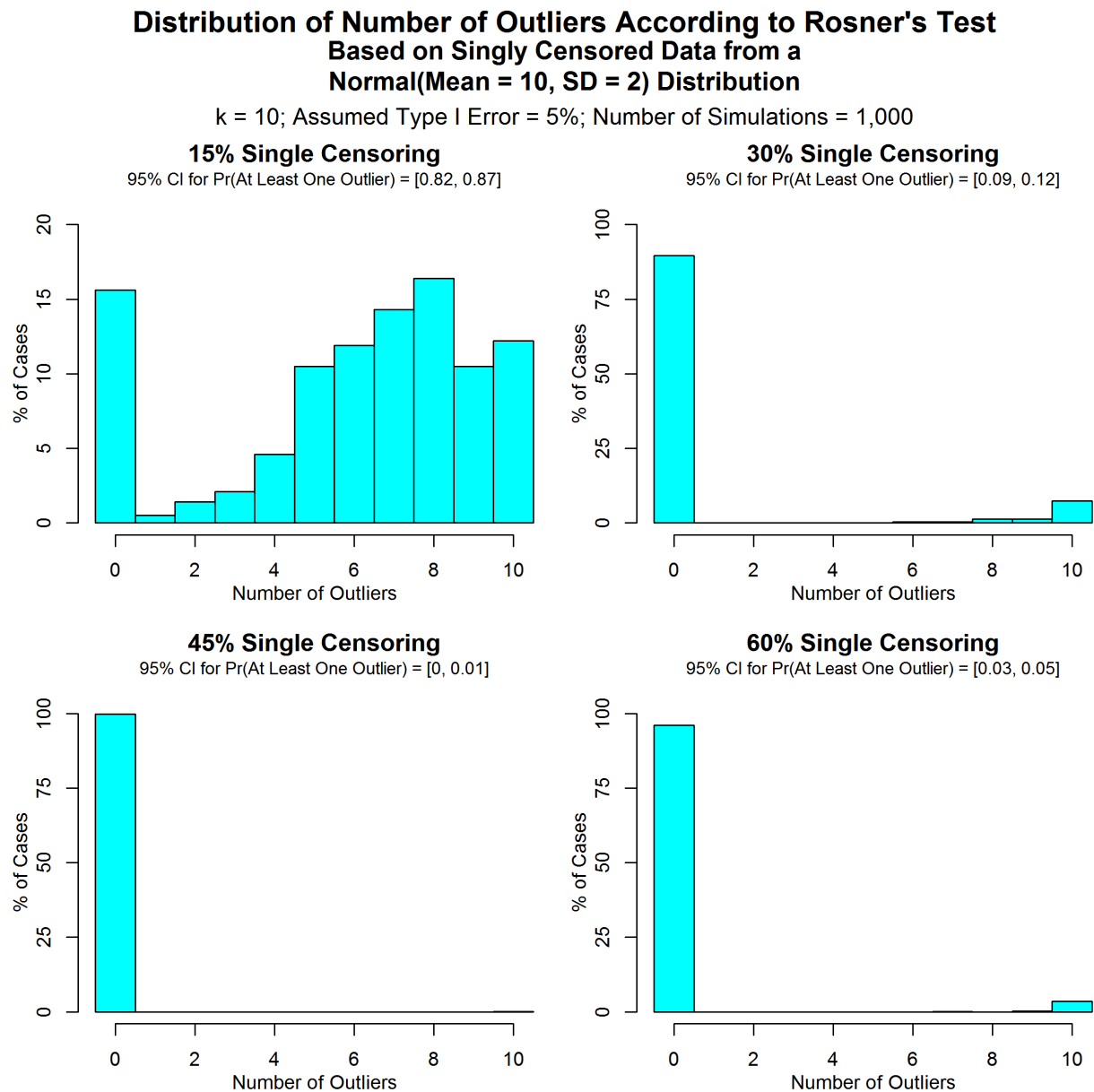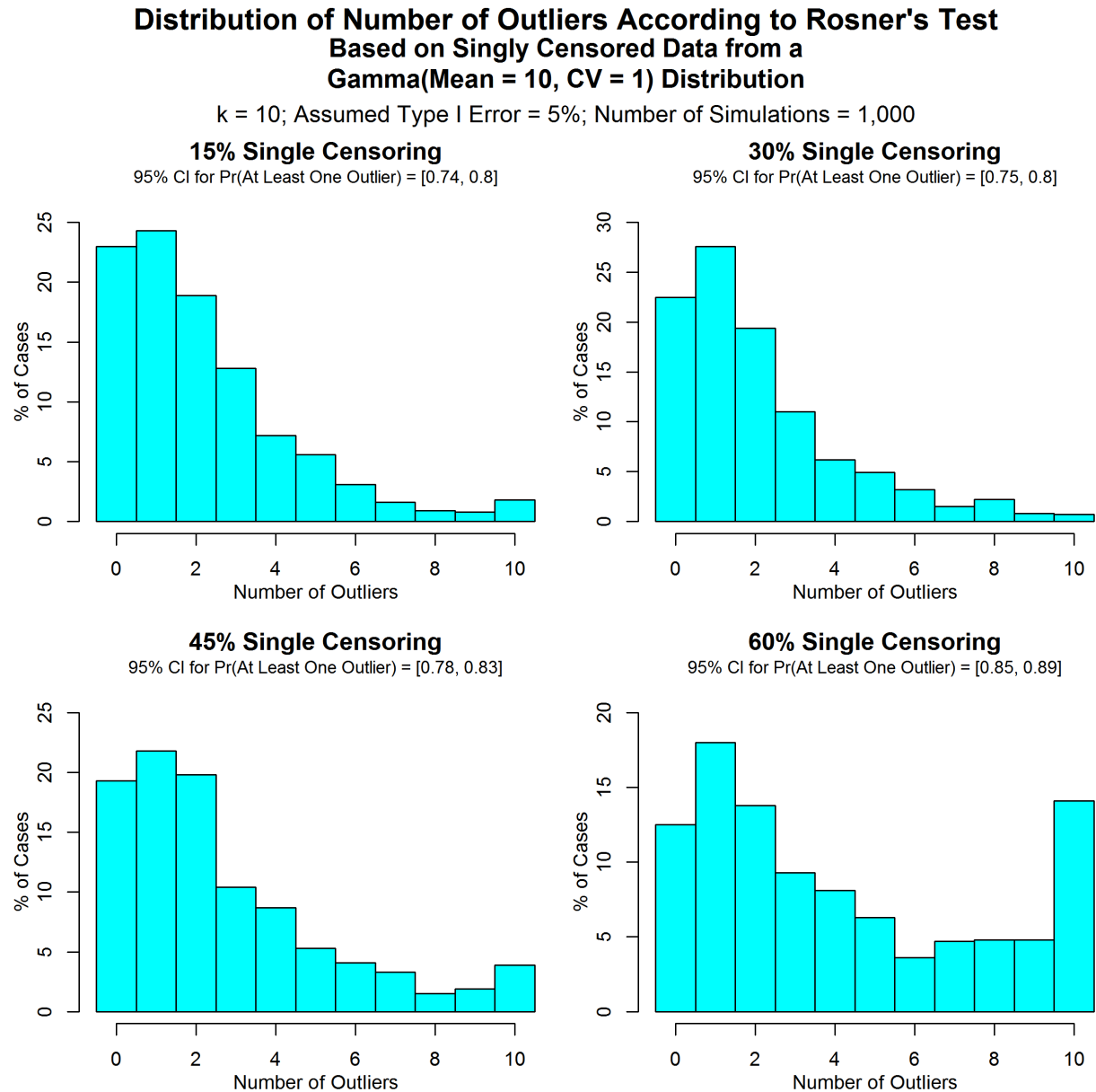95% CI for Pr(At Least One Outlier) = [0.85, 0.89]

Figure 3.3: Distribution of Number of Outliers According to Rosner's Test Based on Lognormal Distribution and Various Censoring Levels

Distribution of Number of Outliers According to Rosner's Test
Based on Singly Censored Data from a
Lognormal(Mean = 10, CV = 1) Distribution
k = 10; Assumed Type I Error = 5%; Number of Simulations = 1,000

15% Single Censoring
95% CI for Pr(At Least One Outlier) = [0.88, 0.92]

30% Single Censoring
95% CI for Pr(At Least One Outlier) = [0.85, 0.89]

45% Single Censoring
95% CI for Pr(At Least One Outlier) = [0.87, 0.91]

60% Single Censoring
95% CI for Pr(At Least One Outlier) = [0.89, 0.93]